University of Alberta

ESTIMATION FOR AGE-PERIOD-COHORT MODELS: WITH APPLICATION TO THE MESOTHELIOMA DATA IN ALBERTA

by

Bei Jiang

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Master of Science**

in

Biostatistics

Department of Mathematical & Statistical Sciences

Edmonton, Alberta
Fall 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-47271-2
Our file Notre référence
ISBN: 978-0-494-47271-2

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Age-Period-Cohort (APC) models have been widely used by epidemiologists to analyze the time trends of incidence and mortality rates of various disease. The linear dependency among three factors: "Cohort=Period−Age" results in the well-known "non-identifiability" problem. In this article, we review the most popular "solutions" that are built on estimable functions–parametrization suggested by Holford (1983) and the overall drift concept proposed by Clayton and Schifflers (1987). Next, we present how regression splines can be used to extend the current APC-model framework by Heuer (1997). Suggestions to practitioners on parameterization of APC models by Carstensen (2007) are also reviewed. Finally we consider the use of smoothing splines in the APC modeling. We show through simulation studies that the model using smoothing splines gives more stable estimates of the estimable functions compared to the other two smoothing techniques. All the methods are illustrated to Alberta Mesothelioma data from 1985 to 2004.

Acknowledgements

I would like to thank all of those who have supported and helped me throughout my studies.

I am deeply indebted to my supervisor Dr. K.C. Carriére. Without her tremendous encouragement, guidance and support this project could not be going so smoothly. Her deep knowledge and great insight on this project saved me from getting stuck in many issues on the way. I am so grateful for the great opportunity to work at Alberta Cancer Board (ACB) she provided me, from which I really learned a lot. I have been feeling so lucky to study under the guidance of Dr. K.C. Carriére and I could not forget her so prompt decision to accept me.

I would like to thank those that helped me at ACB, Munira, Marilyn and Zaman for their detailed introduction and explanation of the Mesothelioma project in every aspect and their knowledge in epidemiology which is really helpful.

I would like to thank all the faculty and staff in the Department of Mathematical and Statistical Sciences for their assistance and for providing such an excellent studying environment. Special thanks to those faculties who have taught me courses for leading me to further explore the wonderful world of statistics.

I am very grateful to my parents for their love and forever support. I cannot express how much you mean to me. I will never forget your sacrifice. To my twin sister for both cheering me up and making me angary and for being present whenever I need her.

Last but not least I would like to thank my husband Linglong Kong. Your support has been a constant source of strength.

Contents

1 Introduction 1

2 Review of the Literature 5

2.1 The Identification Problem 5

2.2 Parameterization and Estimable Functions by Holford (1983) 8

2.3 The Overall Drift and Reduced Two-factor Models 11

2.4 Regression Splines in the Age-Period-Cohort Model 13

2.5 Carstensen's Parameterization 16

2.5.1 Explicit Drift Parameter 17

2.5.2 Fitting Model Sequentially 18

2.6 Splines and Generalized Additive Models 19

2.6.1 Regression Splines 20

2.6.2 Roughness Penalty and Smoothing Splines 23

2.6.3 Additive Models: P-Covariate Case 27

2.6.4 Generalized Additive Models 30

3 Smooth Age-Period-Cohort Models Using Smoothing Splines 32

3.1 Holford's Parameterization to the Smooth APC Models 32

3.2 Smooth APC Models Using Smoothing Splines 33

3.2.1 smoothing splines 34

3.2.2 The Parameterization Using Smoothing Splines 36

4 Simulation Study and Data Analysis 38

4.1 Simulation Study 38

4.2 Analysis of Mesothelioma Data in Alberta 69

5 Conclusion and Discussion 73

Bibliography 75

List of Tables

2.1 Age-by-Period Display 5

2.2 Age-by-Cohort Display 5

2.3 Period-by-Cohort Display 6

2.4 Design Matrix for $I = J = 3$ 9

4.1 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 42

4.2 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 43

4.3 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 44

4.4 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$ 45

4.5 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$ 46

4.6 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$ 47

4.7 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 48

4.8 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 49

4.9 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$ 50

4.10 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 51
4.11 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 52
4.12 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 53
4.13 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$ 54
4.14 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$ 55
4.15 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$ 56
4.16 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 57
4.17 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 58
4.18 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$ 59
4.19 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 60
4.20 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 61
4.21 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 62
4.22 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$ 63
4.23 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$ 64

4.24 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$ 65
4.25 Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 66
4.26 Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 67
4.27 Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$ 68
4.28 Analysis of deviance for the mesothelioma data in Alberta. 69
4.29 Results for the estimable linear trends in the mesothelioma incidence. 70
4.30 Chi-square tests for the estimable curvature effects for mesothelioma data. . . . 70

List of Figures

4.1 Constructed Birth Cohort Effect Curves I, II and III. 39

4.2 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 42

4.3 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 43

4.4 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 44

4.5 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$. 45

4.6 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$. 46

4.7 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$. 47

4.8 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 48

4.9 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 49

4.10 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$. 50

4.11 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$. 51

4.12 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$. 52

4.13 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$. 53

4.14 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$. 54

4.15 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$. 55

4.16 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$. 56

4.17 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$. 57

4.18 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$. 58

4.19 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$. . . . 59

4.20 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = 5.587 \times 10^{-2}$. 60

4.21 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$. . . 61

4.22 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$. . 62

4.23 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$. . . 63

4.24 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$. . . 64

4.25 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$. . 65

4.26 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$. . 66

4.27 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$. . 67

4.28 Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$. 68

4.29 Estimates of curvature effects from three smooth APC models using 3, 3 and 4 Dfs for Age, Period and Cohort respectively. 72

4.30 Estimates of curvature effects from the APC factor model with 5, 4 and 8 Dfs for Age, Period and Cohort respectively. 72

Chapter 1

Introduction

Age-Period-Cohort (APC) models are widely used by epidemiologists to analyze trends in disease incidence and mortality. The purpose of fitting APC models is to decompose the rates into age, period and cohort effects, which are the parameters of interest. The data to be considered are number of cases and population as the denominator for the rates, which can be arranged in a two-way contingency table with one dimension for age groups, indexed by $i = 1, ..., I$, and the other dimension for period groups, indexed by $j = 1, ..., J$. Such a table is also called as the Lexis diagram (Keiding 1990). Due to the linear dependency of the three factors: Cohort=Period−Age, all those in a given age group i at period j correspond to the same birth cohort as those in age group $i+1$ at period $j+1$. Thus, if the age and period groups have equally spaced intervals, the birth cohorts are the diagonals in the Lexis diagram, which will be longer than the interval for age/period groups and overlap to some extent. The birth cohort indexed by $k = 1, ..., K$, can be determined uniquely by the age i and period j,

$$k = j + I - i. \tag{1.1}$$

For each cell (i, j) in the lexis diagram, the number of cases n_{ijk} for Age i, Period j and Cohort $k = j + I - i$ is given and are often assumed to be independently distributed as Poisson with mean μ_{ijk}. The person-years N_{ij} has usually been considered to be fixed and treated as an offset term in the model (Armitage, 1966). It is usually assumed that each factor has an additive effect on the log scale:

$$\log(\mu_{ijk}) \equiv \eta_{ijk} = \mu + \alpha_i + \pi_j + \gamma_k, \tag{1.2}$$

where μ is the intercept and α_i, π_j, γ_k represent the age, period and cohort effects, respectively.

To avoid overparameterization, the usual constraints apply:

$$\sum_i \alpha_i = \sum_j \pi_j = \sum_k \gamma_k = 0 \tag{1.3}$$

Alternatively, constraints can be set by identifying one of each of the age, period, and cohort groups as the reference, which is the default constraint setting in **SAS**. Unfortunately, the model (1.2) is still not identifiable even under the usual constraint (1.3).

Fienberg and Mason(1978) used the log-linear contingency tables to present the identification problem. They viewed an effect as the interaction contrast between two main effects and showed that it is impossible to estimate the linear effect of each of the three factors. They pointed out that only one additional linear constraint, which can be written as $c^T\theta = 0$ where $\theta^T = (\alpha^T, \pi^T, \gamma^T)$ is necessary to ensure identifiability of all the parameters in the model, say $\alpha_1 = 0$ or $\pi_1 = 0$ or $\gamma_1 = 0$. Many authors argued about the choice of c, which should be made on the defensible priori information. Kupper et al.(1983, 1985) considered the commonly used multiple classification APC model in the form of fixed-effect ANOVA-type model to prove that the design matrix is one less than a full-column rank, so one additional constraint is necessary to ensure the identifiability of all the parameters. With constraints imposed, it is well known that the deviance of the model remains the same but the effect estimates can be totally different. This is known as the nonidentifiability problem in age-period-cohort analysis.

There are many "solutions" that aim to handle this nonidentifiability problem without attempting to impose arbitrary constraints. Osmond and Gardner(1982) suggested using the penalty function, which is minimized to derive the necessary extra linear constraint. Robertson and Boyle(1986) attempted to overcome the identifiability problem by using the individual records of cases and forming a three-way table of age group, time period and birth cohort. Lee and Lin(1996) modeled the birth cohort effects as the first-order autoregressive time series.

More recently methods are focused on the utility of estimable functions that are invariant to the selection of constraints on the parameter vector (Holford 1983,1991,1996; Clayton and Schifflers 1987). Their parametrization decompose each factor into a linear effect and curvature effects or deviations from linearity and allows the identifiability of the three curvature effects and the sum of any two linear effects out of the three linear effects. Heuer (1997) introduced

the regression splines, especially Natural Splines, to the Age-Period-Cohort analysis. This approach decomposes the basis of each spline through orthogonal projection into linear and nonlinear terms and the nonidentifiability problem can be handled by excluding one of the two linear effects of period or curvature effect, which in essence are still based on the parametrization by Holford (1983) but provides a framework to include the interactions into the model by using Kronecker product of the spline functions. Using the ideas of both Holford (1983, 2006) and Heuer (1997), Carstensen (2007) gave a set of recommendations for practitioners and introduced the package "Epi" in **R** that implements them.

In practice, we may consider a hierarchy of log-linear models: age, age-drift, age-cohort/age-period and finally age-period-cohort models. Deviance will be used as a measure of "goodness of fit" of the models. When the reduced two-factor model is the final model, special attention should be paid to the interpretation of the effect estimates. Although the third factor is not in the model, one cannot conclude that this factor does not have significant effect. Rather its curvature effect is not significant. This is obvious using the parametrization by Holford (1983). Kupper(1983) demonstrated this both theoretically and by example in the case of Least-square regression.

Further, we consider the use of Generalized Additive Models (Hastie and Tibshirani 1990) for the estimation. The smooth version of the APC model using smoothing splines can be specified as:

$$\log(\mu_{ijk}) = \log(N_{ij}) + \mu + f_a(a_i) + f_p(p_j) + f_c(c_k) \tag{1.4}$$

where a_i and p_j are the actual unaggregated values for age and period in one-year intervals and c_k will be identified as the mean values in the corresponding birth cohorts of two-year intervals through $c_k = p_j - a_i$. N_{ij} is the value of person-years, and $\log(N_{ij})$ is treated as an offset term; $f_a(a_i)$,$f_p(p_j)$ and $f_c(c_k)$ are spline functions for age, period and cohort respectively. Since the cubic smoothing spline is one of the most popular smoothing technique in practice, here we will just focus on the case when cubic smoothing splines are used for all three factors.

A cubic smoothing spline has symmetric smoother matrix with eigenvalues in $(0, 1]$, its smoother matrix can be partitioned into two parts: (1) a projection part–the eigenspace spanned by the eigenvectors for the eigenvalues that equal 1 corresponding to the space spanned by a

constant and linear function of the predictors; and (2) shrinking part–the eigenspace spanned by the eigenvectors for those eigenvalues greater than 0 and strictly less than 1 corresponding to the space spanned by nonlinear functions of the predictors. Hence estimation of the APC additive models fitted by cubic smoothing splines using local scoring algorithm as outer loop and backingfitting algorithm as inner loop can be separated into linear components (corresponding to the projection part) and nonlinear components (corresponding to the shrinking part). The linear dependency among three factors still exists in this framework and can only occur in the projection parts. What is worth noticing is, in the GAM situation, the orthogonal projection is proceeded in the transformed coordinate using certain weight matrix generated in the step of local scoring or in other words, the linear component is a weighted least squares fit of the fitted curve on the predictor and the residual is the nonlinear component. But the nonlinear components are estimable and then conclusions concerning estimable functions by Holford (1983) also hold, so similar parametrization can be conducted and non-identifiability problem will has similar solution in the framework of Generalized Additive Model (GAM) using smoothing splines.

In Chapter 2, we will review the identification problem inherent in age-period-cohort models, the "solution" that are based on the estimable function–parametrization suggested by Holford (1983) and the overall drift concept proposed by Clayton and Schifflers (1987). Next, we will present how to parameterize the age-period-cohort by regression splines as suggested by Heuer (1997) and still based on the parametrization by Holford (1983). A short review of the parametrization by Carstensen (2007) will also be given. Lastly, we reviewed some theories about smoothing splines and GAM that we find useful in explaining the parametrization and proving the estimable functions in the Chapter 3. Chapter 3 will give the similar solution as Holford's (1983) for the APC model using smoothing splines. In Chapter 3, we will use the Holford's and Heuer's approach reviewed in Chapter 2 and the newly considered one in Chapter 3 to analyze the Mesothelioma data in Alberta, 1985 to 2004. In Chapter 4, a brief comparison of those approaches and concluding remarks will be given.

Chapter 2

Review of the Literature

2.1 The Identification Problem

The identification problem is most simply illustrated in the context of the data from three periods of time and divided into three age groups. Corresponding to this structure are five birth cohorts.

The Contingency Table Representation

Fienberg and Mason (1979) presented the three-age and three-period situation in three different cross-classification tables: age-period array in Table 2.1, age-cohort array in Table 2.2 and period-cohort array in Table 2.3. Tables 2.2 and 2.3 are incomplete contingency tables in which the dashes represent structural zeros–categories that are a priori impossible given the way in which the data are collected and restructured.

Table 2.1: Age-by-Period Display

age\period	1	2	3
1	n_{113}	n_{124}	n_{135}
2	n_{212}	n_{223}	n_{234}
3	n_{311}	n_{322}	n_{333}

Table 2.2: Age-by-Cohort Display

age\cohort	1	2	3	4	5
1	–	–	n_{113}	n_{124}	n_{135}
2	–	n_{212}	n_{223}	n_{234}	–
3	n_{311}	n_{322}	n_{333}	–	–

When the effects of age, period and cohort are considered simultaneously, then the third effects are measured in terms of interaction contrasts of the other two effects. That is, in Table 2.1 cohort effects represent the interaction between age and period; in Table 2.2 period

Table 2.3: Period-by-Cohort Display

period\cohort	1	2	3	4	5
1	n_{311}	n_{212}	n_{113}	–	–
2	–	n_{322}	n_{223}	n_{124}	–
3	–	–	n_{333}	n_{234}	n_{135}

effects represent the interaction between age and cohort; and in Table 2.3 age effects represent the interaction between period and cohort. In the following analysis, we use the notation: $\Omega_{ijk} = \log(\mu_{ijk}/N_{ijk})$.

When examining the period-by-cohort array in Table (2.3), we find that there are two nonredundant interaction contrasts, involving 2×2 subtables of adjacent cells, for estimating age effects:

$$\Omega_{113} - \Omega_{212} - \Omega_{223} + \Omega_{322} = \alpha_1 - 2\alpha_2 + \alpha_3 = -3\alpha_2 \tag{2.1}$$

(since $\alpha_1 + \alpha_2 + \alpha_3 = 0$), and

$$\Omega_{124} - \Omega_{223} - \Omega_{234} + \Omega_{333} = \alpha_1 - 2\alpha_2 + \alpha_3 = -3\alpha_2 \tag{2.2}$$

The two interaction contrasts for the log-odds ratios turn out to be the same contrast for the age effect parameters, α_2 and this is the only estimable age effects. Similarly, the two interaction contrasts for the 2×2 subtables of the adjacent cells in the age-by-cohort array both reduce to the same contrast for the period effect parameters, π_2 and this is the only estimable period effect.

Finally, for the age-by-period array, the four interaction contrasts involving 2×2 subtables of adjacent cells are:

$$\Omega_{113} - \Omega_{124} - \Omega_{212} + \Omega_{223} = 2\gamma_3 - \gamma_2 - \gamma_4 \tag{2.3}$$

$$\Omega_{124} - \Omega_{135} - \Omega_{223} + \Omega_{234} = 2\gamma_4 - \gamma_5 - \gamma_3 \tag{2.4}$$

$$\Omega_{212} - \Omega_{223} - \Omega_{311} + \Omega_{322} = 2\gamma_2 - \gamma_1 - \gamma_3 \tag{2.5}$$

$$\Omega_{223} - \Omega_{234} - \Omega_{322} + \Omega_{333} = 2\gamma_3 - \gamma_2 - \gamma_4 \tag{2.6}$$

Since Expressions (2.3) and (2.6) involve the same contrast, there are only three equations can be used to estimate four independent cohort parameters. By taking linear combinations

of expressions of Expressions (2.3), (2.4) and (2.5) and using the constraint $\sum_k \gamma_k = 0$, the cohort effects $\gamma_3, \gamma_1 + \gamma_5$ and $\gamma_1 + 2\gamma_4$ are estimable.

Therefore each type of effects is lacking one estimable parameter. Another way to look at this identification problem is: due to the constraint $\sum_i \alpha_i = 0$, the only estimable parameter α_2 allows us to estimate the quadratic effect for age $\alpha_1 - 2\alpha_2 + \alpha_3$. But we are unable to get the estimate for linear age effect $\alpha_3 - \alpha_1$. Similarly, we can get the the quadratic effect for period $\pi_1 - 2\pi_2 + \pi_3$ but not the linear effect $\pi_3 - \pi_1$. For the cohort effect, from Expression (2.3) to (2.5), we cannot obtain the linear effect. What we have seen here is that when the effects are described into two components: linear trend and nonlinear trend, the nonlinear effects are estimable while the linear effects are not, which is actually the cause of the identification problem.

The identification problem complicates the analysis, but it can be overcome by imposing an *identification specification*; that is, one additional constraint on the effect parameters, such as $\alpha_1 = \alpha_2$, or $\pi_1 = \pi_2$, or $\gamma_1 = \gamma_2$ is sufficient to identify all parameters. The goodness-of-fit of an Age-Period-Cohort model to data can still be tested despite the non-identifiability of certain parameters.

Least-squares Regression representation

The identification problems in APC models can be equivalently seen when working with the model in a least-squares regression form (Kupper, 1985). The constraint (1.1) implies that only the first $(I-1)$ age effects, the first $(J-1)$ period effects and the first $(I+J-2)$ birth cohort effects needed to be estimated. Then, the log-linear model (1.2) has the least-squares regression presentation as:

$$Y_{ij} = \mu + \sum_{i=1}^{I-1} \alpha_i A_i + \sum_{j=1}^{J-1} \pi_j P_j + \sum_{k=1}^{I+J-2} \gamma_k C_k \qquad (2.7)$$

where, Y_{ij} represent the incidence rate or log-transformed rate. A_i, P_j and C_k are the dummy variables for the i^{th} age group, j^{th} period and k^{th} birth cohort, respectively. The subscript k of the cohort effects can be determined uniquely by i and j.

Accordingly, the $(IJ)\times[2(I+J)-3]$ dimensional design matrix $\mathbf{X}$ of the matrix presentation

of the model (2.7) can be defined, namely

$$E(\mathbf{Y}) = \mathbf{X}b \tag{2.8}$$

where the response vector is

$$\mathbf{Y}' = (Y_{11}, ..., Y_{1J}; Y_{21}, ..., Y_{2J}; ...; Y_{I1}, ..., Y_{IJ}) \tag{2.9}$$

and the parameter vector is

$$b' = (\mu; \alpha_1, ..., \alpha_{I-1}; \pi_1, ..., \pi_{J-1}; \gamma_1, ..., \gamma_{I+J-2}) \tag{2.10}$$

The ordinary least-squares estimates $\hat{b}$ is the solution of the normal equation:

$$\mathbf{X}'\mathbf{X}b = \mathbf{X}'\mathbf{Y} \tag{2.11}$$

But because the design matrix $\mathbf{X}$ is one-less than full column rank, the matrix $\mathbf{X}'\mathbf{X}$ is not invertible, and there does not exist a unique solution to (2.11). This is the identification problem of APC analysis. Actually, due to the linear dependence: Cohort=Period $-$Age, some columns of the matrix $\mathbf{X}$ can be expressed linearly by other columns.

Kupper et al. (1985, theorem 3.1) have shown that for columns of the $IJ \times [2I + 2J - 3]$ matrix $\mathbf{X}$, there exists the following linear dependency:

$$\sum_{i=1}^{I-1}[i - \frac{I+1}{2}]A_i - \sum_{j=1}^{J-1}[j - \frac{J+1}{2}]P_j + \sum_{k=1}^{I+J-2}[k - \frac{I+J}{2}]C_k = \mathbf{0} \tag{2.12}$$

The above equation actually reveals that "the linear component of age"$-$"the linear component of period"+"the linear component of cohort"$= 0$, since $[i - \frac{I+1}{2}]$, $[j - \frac{J+1}{2}]$ and $[k - \frac{I+J}{2}]$ are the orthogonal polynomial coefficients for accessing the linear trend of corresponding factors as suggested by Holford (1983). This result is consistent with other authors' conclusion (for example, Fienberg and Mason (1979); Holford (1983)) that the linear effects of age, period and cohort are not individually estimable when the three factors are considered simultaneously.

2.2 Parameterization and Estimable Functions by Holford (1983)

Holford (1983) did not make any attempt to impose any constraints to ensure identifiability but concentrated only on the estimable functions. This approach is adopted later on by Heuer

(1997) in the age-period-cohort modeling using natural splines. Holford (1983) suggested decompose the effects of age, period and cohort into two components: linear trend and curvature (or deviations from linearity).

Table 2.4: Design Matrix for $I = J = 3$

i	j	k	$\mathbf{A}_C$	$\mathbf{P}_C$		$\mathbf{C}_C$		$\mathbf{A}_L$	$\mathbf{P}_L$	$\mathbf{C}_L$
1	1	3	1	1	−2	0	6	−1	−1	0
	2	4	1	−2	−1	−2	−4	−1	0	1
	3	5	1	1	2	1	1	−1	1	2
2	1	2	−2	1	−1	2	−4	0	−1	−1
	2	3	−2	−2	−2	0	6	0	0	0
	3	4	−2	1	−1	−2	−4	0	1	1
3	1	1	1	1	2	−1	1	1	−1	2
	2	2	1	−2	−1	2	−4	1	0	−1
	3	3	1	1	−2	0	6	1	1	0

To illustrate, age effects α_i ($\sum_i \alpha_i = 0$) are considered. The linear trend of age α_L can be described by the linear contrast:

$$\alpha_L = C \sum_i A_L(i)\alpha_i = \frac{\sum_{i=1}^{I}[i - \frac{I+1}{2}]}{\sum_{i=1}^{I}[i - \frac{I+1}{2}]^2}\alpha_i \tag{2.13}$$

where, $A_L(i) = [i - (I+1)/2], i = 1, 2, ..., I$, i.e. the coefficients of the first-order orthogonal polynomials and $C = (\sum_i A_L(i)^2)^{-1}$. The curvature component of age effects $\tilde{\alpha}_i$ is given by the age effects with the liner trend of age removed:

$$\tilde{\alpha}_i = \alpha_i - A_L(i)\alpha_L \tag{2.14}$$

If $A_{Cl}(i)$, $l = 2, ..., I-1$ are the coefficients of the second- and higher-order orthogonal polynomials and α_{Cl}, $l = 2, ..., I-1$ are the corresponding parameters, the curvature component of age effects $\tilde{\alpha}_i$ has an alternative expression:

$$\tilde{\alpha}_i = \sum_{l=2}^{I-1} A_{Cl}(i)\alpha_{Cl} \tag{2.15}$$

Obviously, we have $\sum_i A_L(i)A_{Cl}(i) = 0$, $l = 2, ..., I$ and $\sum_i A_L(i)\tilde{\alpha}_i = 0$. That is, the curvature component $\tilde{\alpha}_i$ generated by the above methods do have the liner trend removed and the linear trend and the curvature component are orthogonal to each other. Similarly, we can

partition the period and cohort effects in the same way,

$$\pi_j = P_L(j)\pi_L + \sum_{l=2}^{J-2} P_{Cl}(j)\pi_{Cl} \tag{2.16}$$

$$\gamma_k = C_L(k)\gamma_L + \sum_{l=2}^{I+J-1} C_{Cl}(k)\gamma_{Cl} \tag{2.17}$$

where $P_L(j) = j - \frac{J+1}{2}$ and $C_L(k) = k - \frac{K+1}{2}$.

Furthermore, we can form the overall design matrix after the parametrization to the model 1.2,

$$\mathbf{X} = (\mathbf{1}, \mathbf{A}_C, \mathbf{P}_C, \mathbf{C}_C, \mathbf{A}_L, \mathbf{P}_L, \mathbf{C}_L) \tag{2.18}$$

where corresponding parameters are $\mathbf{b}' = (\mu, \alpha_C', \pi_C', \gamma_C', \alpha_L, \pi_L, \gamma_L)$. Alternatively, other than using the orthogonal polynomials, $\mathbf{A}_C$, $\mathbf{P}_C$ and $\mathbf{C}_C$ can be found by projecting the design matrix for age, period and cohort to the orthogonal space of the linear vector $\mathbf{A}_L$, $\mathbf{P}_L$ and $\mathbf{C}_L$, respectively. For the model of the form 1.2, the matrix is just columns of indicators for each level of age, period and cohort. Table 2.4 displays the design matrix for the case of $I = J = 3$ using the parametrization by Holford (1983).

Due to the linear dependency of Cohort=Period−Age, we have

$$\mathbf{C}_L = \mathbf{P}_L - \mathbf{A}_L \tag{2.19}$$

Hence the design matrix $\mathbf{X}$ in (2.18) is one less than full column rank and generalized inverse of $\mathbf{X}'\mathbf{X}$ have to be employed to obtain a least-squares solution.

By the definition given by Searle (1971, Ch5.4), the linear function, $\mathbf{q}'\mathbf{b}$ of the parameters is estimable if $\mathbf{q}'\mathbf{b} = \mathbf{t}\log\lambda$ for any $\mathbf{t}$, where $\log\lambda$ is a vector of log-transformed incidence rates. The estimable functions above are invariant to the particular constraint made on the parameters. If the least-squares method is used, the estimable functions are also best linear unbiased estimates (BLUEs). However when using maximum likelihood or iterative proportional fitting method for Poisson random variables, this property does not hold any more. According to Searle (1971, P. 185), it is sufficient to check if $\mathbf{q}'\mathbf{H} = \mathbf{q}$, where $\mathbf{H} = \mathbf{G}\mathbf{X}'\mathbf{X}$, and $\mathbf{G}$ is a generalized inverse of $\mathbf{G}$.

For the design matrix $\mathbf{X}$ in (2.18), we partition $\mathbf{X} = (\mathbf{X_1}|\mathbf{C}_L)$, since now $\mathbf{X_1}$ is of full column rank. Now,

$$\mathbf{X'X} = \begin{bmatrix} \mathbf{X_1}'\mathbf{X_1} & \mathbf{X_1}'\mathbf{C}_L \\ \mathbf{C}_L'\mathbf{X_1} & \mathbf{C}_L'\mathbf{C}_L \end{bmatrix} \tag{2.20}$$

is invertible. When we use the generalized inverse of $\mathbf{X'X}$ as the following:

$$\mathbf{G} = \begin{bmatrix} (\mathbf{X_1}'\mathbf{X_1})^{-1} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \tag{2.21}$$

we have $\mathbf{H} = \mathbf{GX'X}$ as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{I} & (\mathbf{X_1}'\mathbf{X_1})^{-1}\mathbf{X_1}'\mathbf{C}_L \\ \mathbf{0} & 0 \end{bmatrix} \tag{2.22}$$

where the upper right-hand portion of $\mathbf{H}$, i.e., $\mathbf{X_1}'(\mathbf{X_1})^{-1}\mathbf{X_1}'\mathbf{C}_L$ is the least-squares solution $\hat{L}$ to the equation $\mathbf{C}_L = \mathbf{X_1}L$. Using the linear dependency in (2.20), H is reduced to:

$$\mathbf{H} = \begin{bmatrix} & \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ 1 \end{pmatrix} \\ \mathbf{I} & \\ \mathbf{0} & 0 \end{bmatrix} \tag{2.23}$$

The linear effect of age α_L can be expressed as $\mathbf{q'b} = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & 0 \end{pmatrix}\mathbf{b}$ and obviously $\mathbf{q'H} \neq \mathbf{q'}$. Hence α_L is not an estimable function. Similarly, it is easy to see that the linear trend of period and birth cohort, π_L and γ_L, are not estimable, as demonstrated by Fienberg and Mason (1979). However, $\alpha_L + \pi_L$ and $\pi_L + \gamma_L$ are estimable, in other words, the linear effects of the three factors can be not distinguished from each other. In general, any function of the three linear components which has the form of $d_1\alpha_L + d_2\pi_L + (d_2 - d_2)\gamma_L$ with arbitrary d_1 and d_2 is estimable. For curvature components, any linear function given by $\begin{pmatrix} \mathbf{q}_c' & 0 & 0 & 0 \end{pmatrix}\mathbf{b}$ with arbitrary $\mathbf{q}_c$ is estimable, which means that the curvature components of the three factors are all estimable.

2.3 The Overall Drift and Reduced Two-factor Models

Clayton and Schifflers (1987) describe an approach to analyzing time trends that only consider age effect and the overall "drift" with period or cohort effect:

Age+drift with period model:

$$\Omega_{ij} = \mu + \alpha_i + j \cdot \beta_p \tag{2.24}$$

Age+drift with cohort model:

$$\Omega_{ik} = \mu + \alpha_i + k \cdot \beta_c \tag{2.25}$$

These two models are more restrictive than the two-factor models as the period or cohort effect is forced to be linear. Further we cannot distinguish between these two models. Actually, when applying the relationship among the three indices: $k = j - i + I$, model (2.24) can be written as:

$$\Omega_{ik} = \mu + (i - I)\beta_c + \alpha_i + k \cdot \beta_c \tag{2.26}$$

Therefore these two models will give an identical fit to the data and thus identical deviance also due to the problem. As proved by Holford (1985) that the linear effects of age, period and cohort, α_L, π_L and γ_L can not be distinguished and only the estimates of $\alpha_L + \pi_L$ and $\pi_L + \gamma_L$ can be obtained, with which there are two degrees of freedom associated. Clayton and Schifflers's method is to partition the two degrees of freedom into two components. One is the linear age effect and the other is the "drift", which contain both period and cohort effects. The rationale for this approach is to assume age to be the dominant factor and assign one of the linear components to this factor and the time and cohort effects are linked together known as the "drift".

The introduction of the idea of "drift" led Clayton and Schifflers (1987) to suggest a hierarchicy of models:

1. Age
2. Age+Drift
3. a. Age+Period
   b. Age+Cohort
4. Age+Period+Cohort

The comparisons between $3a$ or $3b$ with 4 give the tests for cohort adjusted for period and vice versa. Here, we use the comparison of the model $3a$ with 4 to illustrate. Each effect can decomposed into a linear effect and curvature effects as suggested by Holford (1983); we can

represent age by $\alpha_i = i^*\beta_a + \alpha_{Ci}$, where $i^* = i - (I+1)/2$. Doing the similar partition for period and cohort, the model $3b$ can be rewritten as:

$$log(\lambda_{ijk}) = \mu + i^*\beta_a + \alpha_{Ci} + k^*\beta_c + \gamma_{Ck} \tag{2.27}$$

and model 4 as:

$$\begin{aligned} log(\lambda_{ijk}) &= \mu + i^*\beta_a + \alpha_{Ci} + j^*\beta_p + \pi_{Cj} + k^*\beta_c + \gamma_{Ck} \\ &= \mu + i^*(\beta_a + \beta_p) + \alpha_{Ci} + \pi_{Cj} + k^*(\beta_p + \beta_c) + \gamma_{Ck} \end{aligned} \tag{2.28}$$

because of $j^* = i^* + k^*$. It is already well known that α_{Ci}, π_{Cj} and γ_{Ck} are the curvature effects and are all estimable in model 4. Obviously, comparing the fit of APC model 4 with AC model $3b$ is actually testing whether there is a significant curvature component for the period effects, $H_0 : \pi_{C1} = \pi_{C2} = ... = \pi_{CJ} = 0$. Likewise, the AC model $3b$ cannot escape the bias, either, in which the estimates of the linear effects of age and cohort obtained are really the estimates of $\beta_a + \beta_p$ and $\beta_p + \beta_c$, respectively.

A significant test that compares the fit of model 2 with either model $3a$ or $3b$ is equivalent to a test for the curvature of period or cohort that adjusts for age and the corresponding linear effect of period or cohort.

2.4 Regression Splines in the Age-Period-Cohort Model

In this section we review how regression splines can be applied to fitting the age-period-cohort models as proposed by Heuer (1997). There are two aspects that needed to be considered. One is on the selection of regression splines. The spline curves should be stable in the tails since out there they are based on fewer observations. This is especially the case for the cohort variable as there are far fewer cells for early and later years than for central years in the lexis diagram. The other aspect is on how to handle the nonidentifiability problem sensibly.

In terms of the first concern, the natural spline known as stable in the tails is recommended by Heuer (1997), which has continuous first and second derivatives and is linear past its most left and right inner knots until the boundary. It has been proven to be sufficiently smooth and flexible in practice. From now on, we will concentrate on natural spline.

Similar to the parametrization by Holford (1983) described in the section 2.2.1, the natural splines for age, period and cohort are to be partitioned into linear and curvature components by

the orthogonal projection. The approach of a spline parametrization on the age-period-cohort model can be implemented as follows:

1. Set up the model matrices for age, period and cohort: $\mathbf{M}_a$, $\mathbf{M}_p$ and $\mathbf{M}_c$, all including the intercept term. For the spline model, they are just the columns of basis vectors for each of the splines,

   - age: $\mathbf{M}_a := [B_{-1}^{age}(a), \cdots, B_{m_a-2}^{age}(a)]$
   - period: $\mathbf{M}_p := [B_{-1}^{per}(p), \cdots, B_{m_p-2}^{per}(p)]$
   - cohort: $\mathbf{M}_c := [B_{-1}^{age}(c), \cdots, B_{m_c-2}^{coh}(c)]$

   where, m_a, m_p and m_c are the number of inner knots for age, period and cohort. Since the recursive definition of the cubic B-spline basis functions are very complicated and are hardly revealing, here we skip the specific definition.

2. Extracting the linear trend from $\mathbf{M}_a$, $\mathbf{M}_p$ and $\mathbf{M}_c$ by projecting each column onto the orthogonal complement of the vector of ones $\mathbf{1}$ and the linear vector of a, p and c, respectively. For age, the corresponding projectors are $(I - J'(J'J)^{-1}J')$ and $(I - a'(a'a)^{-1}a')$, respectively, where I is the identity matrix and J the matrix containing only ones. The natural spline basis vectors in $\mathbf{M}_a$ can then be transformed into:

$$\tilde{B}_i^{age}(a) := (I - J'(J'J)^{-1}J')(I - a'(a'a)^{-1}a')B_i^{age}(a)$$

   with $i = -1, 2, \cdots, m_a - 2$. The resulting matrix $\tilde{\mathbf{M}}_a$ has two fewer columns, i.e., two basis vectors can be replaced by the constant $\mathbf{1}$ and the linear term a. The first and the last basis vectors are appropriate for this reduction since they represent the linear influences for the tails. The final basis vectors for the later spline parametrization would be:

$$[\mathbf{1}, a, \tilde{B}_0^{age}(a), \cdots, \tilde{B}_{m_a-3}^{age}(a)]$$

   We define the nonlinear Natural Spline basis vector for age as:

$$\tilde{\mathbf{M}}_a := [\tilde{B}_0^{age}(a), \cdots, \tilde{B}_{m_a-3}^{age}(a)]]$$

   Similarly, we can do the same partition on the corresponding matrices for period and cohort and then obtain $\tilde{\mathbf{M}}_p$ and $\tilde{\mathbf{M}}_c$.

3. Define the linear components for age, period and cohort by,

$$A_L(a) := a - \frac{S+1}{2}, P_L(p) := p - \frac{T+1}{2}, C_L(c) := c - \frac{S+T}{2}$$

They are orthogonal to the constant term. Here we use S, T to denote the total age groups and period groups to accommodate the case of one-year categories. The log-linear age-period-cohort model under spline parametrization by Heuer has the following form:

$$log(\lambda_{apc}) = \mu + A_L(a)\alpha_L + P_L(p)\pi_L + f^{age}(a) + f^{per}(p) + f^{coh}(c) \tag{2.29}$$

where

$$f^{age}(a) = \tilde{\mathbf{M}}_a\vartheta^a, f^{per}(p) = \tilde{\mathbf{M}}_p\vartheta^p, f^{coh}(c) = \tilde{\mathbf{M}}_c\vartheta^c$$

α_L and π_L are the linear trend parameters for age and period, $\vartheta^a, \vartheta^p, \vartheta^c$ are the parameter vectors for the nonlinear natural spline basis vectors for age, period and cohort respectively.

Since in this model 2.29 only two linear effects are considered, all the parameters are identifiable and hence no identification problem exists anymore.

Interactions between age and period or age and cohort can be modeled easily by modifying 2.29 a little bit. For example, the cohort function $f^{coh}(c)$ in 2.29 can be viewed as a special kind of age vs. period interaction, which can be replaced by a more general form of interaction– tensor product of the Natural Splines for age and period. This age-period-interaction model can be expressed as,

$$log(\lambda_{ap}) = \mu + A_L(a)\alpha_L + P_L(p)\pi_L + f^{age}(a) + f^{per}(p) + f^{age,per}(a,p) \tag{2.30}$$

the two dimensional function for the general form of age-period interaction is given by,

$$f^{age,per}(a,p) := Z^{age,per}(a,p)\varrho \tag{2.31}$$

where ϱ is the coefficient vector with appropriate dimension and

$$Z^{age,per}(a,p) := [A_L(a) * P_L(p), A_L(a) * \tilde{\mathbf{M}}_p, \tilde{\mathbf{M}}_a * P_L(p), \tilde{\mathbf{M}}_a \otimes \tilde{\mathbf{M}}_p] \tag{2.32}$$

with

$$\tilde{\mathbf{M}}_a \otimes \tilde{\mathbf{M}}_p := [\tilde{B}_0^{age}(a) * \tilde{B}_0^{per}(p), \ldots, \tilde{B}_{m_a-3}^{age}(a) * \tilde{B}_{m_p-3}^{per}(p)] \tag{2.33}$$

where $\otimes$ denotes the elementwise multiplication of the row vectors of $\tilde{\mathbf{M}}_a$ and $\tilde{\mathbf{M}}_p$.

That is, the interaction term $f^{age,per}(a,p)$ is the tensor product of the Natural Spline basis for age $[A_L(a), \tilde{\mathbf{M}}_a]$ and the basis for period $[P_L(p), \tilde{\mathbf{M}}_p]$.

Likewise, the age-cohort interaction of this kind can be found and modeled similarly.

2.5 Carstensen's Parameterization

The

general form of the multiplicative age-period-cohort model in Carstensen's Parametrization has the following form:

$$log[\lambda(a,p)] = f(a) + g(p) + h(c) \tag{2.34}$$

where, a, p and $c(= p - a)$ represent the mean age, period and cohort for each cell in the Lexis diagram and are continuous variables. Then the model can predict the rates at any point.

"Usually we have strong evidence to believe that age is the most important factor while period is the least important." So Carstensen's suggested the parametrization based on the following assumptions:

- The age-function should be interpretable as log of age-specific rates in a reference cohort c_0 adjusted for period effects.
- The cohort-function is 0 at the reference cohort c_0, interpretable as log of relative risk relative to cohort c_0.
- Period effect is "0" on average and with "0" slope.

In practice, the above suggestions can be implemented as follows:

1. Fit the model 2.34 using any parametrization to obtain $\hat{f}(a)$, $\hat{g}(p)$ and $\hat{h}(c)$. If **SAS** or **R** are employed, we can just use the default parametrization.
2. Regress $\hat{g(p)}$ on p, i.e., $\hat{g(p)} = \mu_p + \delta_p p + \tilde{g}(p)$.
3. Report the effects as:

   Age-specific incidence rates in cohort c_0: $exp[\hat{f}(a) + \mu_p + \delta_p a + \hat{h}(c_0) + \delta_p c_0]$

Rate-ratio relative to cohort c_0: $exp[\hat{h}(c) - \hat{h}(c_0) + \delta_p(c - c_0)]$

Rate-ratio by period controlled for age and cohort: $exp[\tilde{g}(p)]$

The formation behind this is to parameterize the model 2.34 in the following way:

$$\begin{aligned} log[\lambda(a,p)] &= \hat{f(a)} + \hat{g}(p) + \hat{h}(c) \\ &= \tilde{f}(a) + \tilde{g}(p) + \tilde{h}(c) \end{aligned} \tag{2.35}$$

where

$$\begin{aligned} \tilde{f}(a) &= \hat{f}(a) + \mu_p + \delta_p a + \hat{h}(c_0) + \delta_p c_0 \\ \tilde{g}(p) &= \hat{g}(p) - \mu_p - \delta_p p \\ \tilde{h}(c) &= \hat{h}(c) + \delta_p(c - c_0) - \hat{h}(c_0) \end{aligned} \tag{2.36}$$

and the resulting functions $\tilde{f}(a)$, $\tilde{g}(p)$ and $\tilde{h}(c)$ satisfy the desired constraint, i.e., $\tilde{g}(p)$ is 0 on average and "detrended", $\tilde{h}(c_0) = 0$ and $\tilde{f}(a)$ is the log rates in cohort c_0 when controlled for period.

If we want the period function to be zero at a reference period p_0, interpretable as the log rate ratio relative to period p_0, we just use $\mu_p = \hat{g}(p_0) - \delta_p p_0$, where δ_p is the estimated slope from the regression of $\hat{g}(p)$ on p in 2.36.

This method is based on the parametrization of Holford (1983) and similar to the spline parametrization by Heuer (1997) but the estimated functions have more sensible interpretation. Carstensen (2007) suggested that this can be considered as the solution to the nonidentifiability problem. Next, we will list two easily-implemented variants of the above approach.

2.5.1 Explicit Drift Parameter

We can extract the overall drift (Clayton and Schifflers 1987), include it into the model as a separate parameter and include the "de-trended" period and cohort effects. This corresponds to the following partition of the model 2.34,

$$\begin{aligned} log[\lambda(a,p)] &= \hat{f(a)} + \hat{g}(p) + \hat{h}(c) \\ &= \tilde{f}_c(a) + \delta(c - c_0) + \tilde{g}(p) + \tilde{h}(c) \\ &= \tilde{f}_p(a) + \delta(p - p_0) + \tilde{g}(p) + \tilde{h}(c) \end{aligned} \tag{2.37}$$

where period and cohort functions $\tilde{g}(p)$ and $\tilde{h}(c)$ are "de-trended", i.e., have 0 slope and 0 on average:

$$\begin{aligned} \tilde{g}(p) &= \hat{g}(p) - \mu_p - \delta_p p \\ \tilde{h}(c) &= \hat{h}(c) - \mu_c - \delta_c c \end{aligned} \tag{2.38}$$

with the overall drift defined as: $\delta = \delta_p + \delta_c$

Then the two age functions $\tilde{f}_c(a)$ and $\tilde{f}_p(a)$ have the interpretation of log age-specific rates in the reference period p_0 and c_0, respectively and they can be obtained by,

$$\begin{aligned} \tilde{f}_c(a) &= \tilde{f}(a) + \mu_a + \mu_p + \mu_c + (\delta_a + \delta_p)a + (\delta_p + \delta_c)c_0 \\ &= \hat{f}(a) + \mu_p + \mu_c + \delta_p a + \delta c_0 \\ \tilde{f}_p(a) &= \tilde{f}(a) + \mu_a + \mu_p + \mu_c + (\delta_a - \delta_c)a + (\delta_p + \delta_c)p_0 \\ &= \hat{f}(a) + \mu_p + \mu_c - \delta_c a + \delta p_0 \end{aligned} \tag{2.39}$$

where $\tilde{f}(a)$ is "detrended", i.e., $\tilde{f}(a) = \hat{f}(a) - \mu_a - \delta_a a$

Thus, age-specific rates can be reported referring to either a specific cohort (longitudinal rates) or a specific period (cross=sectional rates). Note that $\tilde{f}_c(a) = \tilde{f}_p(a) + \delta(a - (p_0 - c_0))$, so if there is a positive overall drift ($\delta > 0$) the cohort (longitudinal) age-curve will be steeper than the period (cross-sectional) age curve.

2.5.2 Fitting Model Sequentially

Using a small trick, we can obtain an approximation to the parametrization described above. We fit the age-cohort model first:

$$log[\lambda(a,c)] = f(a) + h(c)$$

In terms of the estimated functions $\hat{f}(a)$ and $\hat{h}(c)$, we can parameterize the above model in the following way,

$$\begin{aligned} log[\lambda(a,c)] &= \hat{f}(a) + \hat{h}(c) \\ &= [\hat{f}(a) - \hat{h}(c_0)] + [\hat{h}(c) - \hat{h}(c_0)] \\ &:= \hat{\tilde{f}}(a) + \hat{\tilde{h}}(c) \end{aligned} \tag{2.40}$$

This will make $\hat{\tilde{f}}(a)$ interpretable as the log age-specific rates in cohort c_0 and $\hat{\tilde{h}}(c)$ as the log rate ratio of cohort c compared to cohort c_0.

Next, the estimates $\hat{\hat{f}}(a)$ and $\hat{\hat{h}}(c)$ from the model 2.40 is then modeled as an offset in the following model with period effects:

$$log[\lambda(a,p)] = \hat{\hat{f}}(a) + \hat{\hat{h}}(c) + g(p) \tag{2.41}$$

The period effects from this model can be interpretable as the residual log rate ratio by period. The estimates obtained by this sequential procedure are not the ML-estimates from fitting age-period-cohort models, they are marginal age-cohort estimates and period estimates conditional on the estimates from the age-cohort model.

If there exists a priori knowledge that cohort effects are the major cause of the change in incidence rates, then the above sequential procedure would be the best way to model the rates, since the period effects are only considered as the residuals conditional on the estimated age and cohort effects.

The above trick can be extended to the case when considering a sequence of log-linear models: first fit the age-drift model and then sequentially add cohort and period effects as "residuals",

1. Age+Drift: $log[\lambda(a,p)] = f(a) + \delta(c - c_0)$
2. Cohort | Age+Drift: $log[\lambda(a,p)] = \hat{f}(a) + \hat{\delta}(c - c_0) + h(c)$
3. Period | Age+Drift+Cohort: $log[\lambda(a,p)] = \hat{f}(a) + \hat{\delta}(c - c_0) + \hat{h}(c) + g(p)$

2.6 Splines and Generalized Additive Models

In this section, we review some aspects of smoothing splines and Generalized Additive Models (GAMs) that are relevant and useful in explaining how the nonidentifiability problem can be dealt with when smoothing splines are used on each of the factors. The model would be then fitted in the framework of generalized additive models.

2.6.1 Regression Splines

Suppose we have data (y_i, x_i), $i = 1, 2, ..., n$ where $l < x_1 < x_2 < ... < x_n < u$. Consider the regression model:

$$y_i = f(x_i) + \epsilon_i \tag{2.42}$$

where ϵ_i's are independently distributed $N(0, \sigma^2)$. Estimate of f can be obtained by minimizing the residual sum of squares:

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 \tag{2.43}$$

Instead of minimizing over all linear linear functions (i.e., functions of the form $f(x) = \beta_0 + \beta_1 x$), we could minimize 2.43 over a larger class of smooth functions, say splines. This is basic idea of using splines in the regression analysis and in this context, splines have usually been referred to as regression splines. Next we will give more details about splines.

Given a set of inner knots denoted by $\xi_1 < \xi_2 < \cdots < \xi_K$ contained in (l, u), splines are piecewise polynomials of degree d that are joined smoothly at those K inner knots. More specifically any such piecewise polynomial denoted as $f(x)$ satisfies the following three conditions:

- On each of the intervals $(l, \xi_1), (\xi_2, \xi_3), \cdots, (\xi_K, u)$, $f(x)$ is $d-$degree polynomial.
- Its $(d-1)^{th}$ derivative $f^{(d-1)}(x)$ is continuous over the whole range of (l, u).
- Its d^{th} derivative $f^{(d)}(x)$ is a step function with jumps at $\xi_1, \xi_2, \cdots, \xi_K$.

All possible such piecewise polynomials span a linear space called *spline space* (de Boor, 1978). Totally we need $d + K + 1$ basis functions to span the spline space, since $d + 1$ basis functions are needed to specify the $d-$degree polynomial for each of the $K + 1$ intervals $(\xi_0 \xi_1], [\xi_2, \xi_3], \cdots, [\xi_{K-1}, \xi_K], [\xi_K, u]$ and 1 additional function for each of the K knots are needed due to the continuity constrain. The spline space then can be spanned by these $d+K+1$ basis functions denoted as $B_1(x), B_2(x), ..., B_{d+K+1}(x)$. Equivalently, any function $f(x)$ in the spline space can be expressed as:

$$f(x) = \sum_{k=1}^{d+K+1} \beta_k B_k(x) \tag{2.44}$$

Fitting the data (y_i, x_i), $i = 1, 2, ..., n$ using splines would be obtained through the multiple least-squares regression of y_i's on the $d+K+1$ basis functions evaluated at the n unique values of x_i's. The estimated fit $\hat{\mathbf{f}} := (\hat{f}(x_1), f(x_2), \cdots, f(x_n))^T$ can be then given by:

$$\hat{\mathbf{f}} = \mathbf{B}\hat{\beta} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{Y} \tag{2.45}$$

where

$$\mathbf{B} = \begin{bmatrix} B_1(x_1) & \cdots & B_{d+K+1}(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \cdots & B_{d+K+1}(x_n) \end{bmatrix} \tag{2.46}$$

One simple choice is to generate the splines by *Truncated Power basis* and hence we get Truncated Power splines (abbreviated as TP-splines), which have the following parametric expression:

$$f(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_d x^d + \sum_{j=1}^{K} \theta_j (x - \xi_j)_+^d \tag{2.47}$$

Where $(x)_+ = x\mathbf{1}(x > 0)$ with $\mathbf{1}$ as the indicator function.

For example, a cubic TP-spline with those K knots at $\xi_1 < \xi_2 < \cdots < \xi_K$ can be represented by a linear combination of the basis functions: $\{1, x, x^2, x^3, (x-\xi_1)_+^3, \cdots, (x-\xi_K)_+^3\}$. These $K+4$ basis functions evaluated at the n observed $x_i's$ will constitute the $n \times (K+4)$ design matrix and the estimates of the $K+4$ parameters can be obtained through ordinary least-squares estimation.

Despite the simple closed form expression of TP-splines, it is sometimes not recommended in practice because large number of knots or too close positions of knots may result in highly correlated basis functions.

Another commonly used spline basis is B-spline basis, which leads to more stable spline curves than TP-spline curves. For the fixed K interior knots at $\xi_1 < ... < \xi_K$, additional $2d+2$ knots outside the range of (ξ_1, ξ_K) are needed to construct the $d+K+1$ B-spline basis functions denoted as: $\{B_{-d,d}(x), B_{-d+1,d}(x), ..., B_{0,d}(x), B_{1,d}(x), ..., B_{K,d}\}$. The full sequence of knots for B-spline is defined at:

$$\xi_{-d} = ... = \xi_0 = l < \xi_1 < ... < \xi_K < u = \xi_{K+1} = ... = \xi_{d+K+1}$$

Then the B-spline basis $\{B_{i,d}, i = -d, ..., K\}$ can be recursively derived as follows:

Let $B_{0,0}, B_{1,0}, ..., B_{K,0}$ be the B-spline basis of degree 0:

$$B_{k,0}(x) = \mathbf{1}x \in [\xi_k, \xi_{k+1}) \tag{2.48}$$

Then for B-splines of degree $d \geqslant 1$ the basis functions can be derived recursively as:

$$B_{k,d}(x) = \frac{x - \xi_k}{\xi_{k+d} - \xi_k} B_{k,d-1}(x) + \frac{\xi_{k+d+1} - x}{\xi_{k+d+1} - \xi_{k+1}} B_{k+1,d-1}(x) \tag{2.49}$$

It can be shown that the basis functions of B-splines and TP-splines span the same spline space despite the additional $2d + 2$ knots defined for the B-spline basis, which are just the boundary points of the observation and hence do not have effect on the shape of the spline curve; and furthermore the two equivalent bases can be transformed from one to the other through a linear transformation, which is omitted here and can be found in (de Boor, 1978).

Further, it can be shown that the B-spline basis functions have the following properties:

- $B_{i,d} = 0$, for $x \notin [\xi_i, \xi_{i+d+1}]$
- $B_{i,d} > 0$, for $x \in [\xi_i, \xi_{i+d+1}]$
- $\sum_{i=-d}^{K} B_{i,d}(x)\theta_i = 1$ for $x \in (l, u)$

The first two properties of B-spline basis will lead to a banded regression matrix and hence more stable estimation of spline curves than TP-spline basis. Further, it can be shown that the constant 1 and x are both in the span of the B-spline basis functions.

A variant of splines are the natural splines, which can be defined for any piecewise polynomials of odd degree. In order to make a spline to be a natural spline, additional constraints that the spline curve is linear beyond the boundary knots (i.e., linear over (l, ξ_1) and $\xi_K, u)$) are needed. This constraint will save four degrees of freedom. That is, a natural cubic spline with K inner knots can be represented by K basis functions and hence only K parameters need to be estimated. Natural splines can be defined for the splines generated by both Truncated Power basis and B-spline basis, which are more stable than their original curves at the boundaries since at that two regions there are usually less data. Since it is hard for our eyes to pick up third or higher order discontinuity, cubic splines or cubic natural splines generated by either B-spline basis or Truncated Power spline basis are believed to be flexible

and smooth enough to satisfy most practical application. The choice of number and position of the knots has a strong influence on the fit of the spline curves and hence should be paid careful attention to get desirable fit. One widely used approach to deal with this is to use a maximal number of knots and a penalty term to control the roughness or wiliness of the spline curves. The smoothing splines to be discussed in the next section just emerge in the context of using the roughness penalty approach to spline curve estimation.

2.6.2 Roughness Penalty and Smoothing Splines

Consider the regression model (2.42) once again: $y_i = f(x_i)+\epsilon_i$. Instead of using least-squares method (i.e., to minimize residual sum of squares $\sum_{i=1}^{n}\{y_i - f(x_i)\}^2$) to get the estimate of $f(x)$, penalty roughness approach is to consider minimizing the the following penalized residual sums of squares:

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda J(f), \tag{2.50}$$

where $J(f)$ is a roughness penalty defined to regulate and control the smoothness of the estimation curves. Here we will focus on the special case with $\lambda J(f) := \int\{f''(t)\}^2 dt$.

Consider the following minimization problem among all functions $f(\cdot)$ with continuous first and second derivatives, find one that minimizes the penalized residual sum of squares:

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda \int\{f''(t)\}^2 dt, \tag{2.51}$$

where λ is a fixed constant called *smoothing parameter* and controls the smoothness of the spline. When the value of λ taken goes from 0 to ∞, $f(\cdot)$ would go from any function that interpolates the data to the simple ordinary least square linear fit, i.e., a larger penalty λ corresponding to a smoother fit while a smaller penalty leading to a rougher fit.

Green and Silverman (1994) showed that among all spline curves $f(\cdot)$ that are twice continuously differentiable and interpolates the points (z_i, x_i) at any given values $z_1, z_2, \cdots, z_n$ the cubic natural spline with knots chosen at all unique values of $x_i's$ turns out to be the solution to the minimization problem 2.51. In that case, the first term in 2.51 is fixed for any $f(\cdot)$ that interpolates the data (i.e., which has the same fitted value at $x_i's$) and the natural spline will minimize the penalty term $\int\{f''(t)\}^2 dt$.

Suppose we have n unique values of $x_i's$ and let $\{\mathbf{N}\}_{ij} = N_j(x_i)$ be the corresponding $n \times n$ nonsingular natural-spline basis matrix, we can write $f(\cdot)$ as

$$f(x) = \sum_{j=1}^{n} N_j(x)\theta_j, \tag{2.52}$$

where the $N_j(x)$ is the j^{th} basis function evaluated at x. It can be easily shown that 2.51 can be expressed in the following form:

$$\text{RSS}(\theta, \lambda) = (\mathbf{y} - \mathbf{N}\theta)^T(\mathbf{y} - \mathbf{N}\theta) + \lambda\theta^T\mathbf{\Omega}_N\theta, \tag{2.53}$$

where $\{\mathbf{\Omega}_N\}_{jk} = \int \mathbf{N}_j''(t)\mathbf{N}_k''(t)dt$. Then the penalized least square estimate of θ is:

$$\hat{\theta} = (\mathbf{N}^T\mathbf{N} + \lambda\mathbf{\Omega}_N)^{-1}\mathbf{N}^T\mathbf{y}, \tag{2.54}$$

which has the form of a generalized ridge regression estimate. We denote the n-dimensional vector of n fitted values of $\hat{f}(x_i)$ by $\hat{f}$ and it can be expressed as:

$$\begin{aligned} \hat{\mathbf{f}} &= \mathbf{N}(\mathbf{N}^T\mathbf{N} + \lambda\mathbf{\Omega}_N)^{-1}\mathbf{N}^T\mathbf{y} \\ &= \mathbf{S}_\lambda\mathbf{y} \end{aligned} \tag{2.55}$$

where $\mathbf{S}_\lambda$ known as the smoother matrix is actually a linear operator: $\mathbf{S}_\lambda : \Re^n \mapsto \Re^n$. From now on $\mathbf{S}_\lambda$ will be referred to the smoother matrix of cubic smoothing splines. Since $\mathbf{N}$ is constructed as a natural cubic spline at unique values of $x_i's$, the smoother matrix $\mathbf{S}_\lambda$ only depends on $x_1, x_2, \cdots, x_n$ and the smoother parameter λ but not on $\mathbf{y}$.

The penalized residual sum of squares 2.51 can be reexpressed in the following form:

$$\text{RSS}(\theta, \lambda) = (\mathbf{y} - \mathbf{f})^T(\mathbf{y} - \mathbf{f}) + \lambda\mathbf{f}^T\mathbf{K}\mathbf{f}, \tag{2.56}$$

where $\mathbf{K} = \mathbf{N}^{-T}\mathbf{\Omega}_N\mathbf{N}^{-1}$. The solution to this minimization problem leads to the *Reinsch* form of the smoothing spline:

$$\mathbf{S}_\lambda = (\mathbf{I} + \lambda\mathbf{K})^{-1} \tag{2.57}$$

Projection and Shrinking

The usual least-squares hat matrix is also a kind of linear operators like the smoother matrix of smoothing splines. Let $\mathbf{B}_\xi$ be the $n \times (K+4)$ design matrix of $(K+4)$ cubic spline basis

functions evaluated at the n values of $x_i's$ with knots $\xi_1, \cdots, \xi_K$ and $K << n$. Then the vector of n fitted values by least-squares is given by:

$$\begin{aligned} \hat{\mathbf{f}} &= \mathbf{B}_\lambda(\mathbf{B}_\lambda^T\mathbf{B}_\lambda)^{-1}\mathbf{B}_\lambda^T\mathbf{y} \\ &= \mathbf{H}_\xi\mathbf{y} \end{aligned} \tag{2.58}$$

Here the linear operator $\mathbf{H}_\xi$ is a projection operator, known as the hat matrix in linear regression. The differences between smoother matrices $\mathbf{H}_\xi$ for regression splines and smoothing splines $\mathbf{S}_\lambda$:

- Both $\mathbf{H}_\xi$ and $\mathbf{S}_\lambda$ are symmetric and positive semidefinite.
- $\mathbf{H}_\xi$ is idempotent (i.e., $\mathbf{H}_\xi\mathbf{H}_\xi = \mathbf{H}_\xi$) while $\mathbf{S}_\lambda\mathbf{S}_\lambda \preccurlyeq \mathbf{S}_\lambda$. In other words, for the projection matrix such as $\mathbf{H}_\xi$, it has K eigenvalues equal to 1, and the rest $n - K$ are all 0 while for the smoother matrix such as the cubic smoothing spline $\mathbf{S}_\lambda$, the first two largest eigenvalues of it are 1, and all the left are ≥ 0. That is why smoothing splines are referred to as *shrinking* smoothers while projection splines are referred to as *shrinking* smoothers.
- $\text{rank}(\mathbf{H}_\xi) = K$ while $\text{rank}(\mathbf{S}_\lambda) = n$

where $K\{= \text{trace}(\mathbf{H}_\xi) = \text{rank}(\mathbf{H}_\xi)\}$ gives the dimension of the projection space (spanned by), the number of basis functions and the number of parameters to be estimated. Similarly in the smoothing spline, the *effective degrees of freedom* is defined as:

$$\text{df}_\lambda = \text{trace}(\mathbf{S}_\lambda) \tag{2.59}$$

Eigen-decomposition of the Smoother Matrix

$\mathbf{S}_\lambda$ is a $n \times n$ and positive semidefinite matrix due to the generation of Natural cubic spline, and hence can be eigenvalue decomposed.

Let $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ be the orthonormal basis of eigenvectors of $\mathbf{S}_\lambda$ with associated eigenvalues $\theta_1 \geq \theta_2, \cdots, \theta_n$, the eigenvalue-decomposition of $\mathbf{S}_\lambda$ is:

$$\mathbf{S}_\lambda = \sum_{k=1}^{n} \theta_k \mathbf{u}_k \mathbf{u}_k^T \tag{2.60}$$

with

$$\theta_k = \frac{1}{1+\lambda d_k}, k = 1, 2, \cdots, n \tag{2.61}$$

where $d_k's$ are the corresponding eigenvalues of $\mathbf{K}$, which is shown to be a nonnegative definite matrix (Golub and van Loan 1983). The followings are true:

- The eigenvectors of $\mathbf{S}_\lambda$ are not affected by the value of λ. So for a particular sequence of $x_i's$, the eigenvectors are the same despite of λ.
- $\text{trace}(\mathbf{S}_\lambda) = \Sigma_{k=1}^{n}\theta_k$. For the projection smoother, however, $\text{trace}(\mathbf{H}_\xi) = \text{rank}(\mathbf{H}_\xi) = K$, which equals the dimension of the projection subspace.

Define $\mathbf{U} = [\mathbf{u}_1, \cdots, \mathbf{u}_n]$ and $\mathbf{D} = \text{diag}[d_1, d_2, \cdots, d_n]$ with $d_k \geq 0$ for $k = 1, \cdots, n$. Since the matrix of eigenvectors $\mathbf{U}$ spans the space of $\mathbf{S}_\lambda$, $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$ can be written as $\hat{\mathbf{f}} = \mathbf{U}\beta$ where β is a vector of length n. Accordingly the eigenvalue decomposition of $\mathbf{K}$ in 2.56 can be expressed as $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^T$. Then the penalized term in 2.56 can be written as $\lambda\beta^T\mathbf{D}\beta$. In this case the smoothing spline $\mathbf{S}_\lambda$ is the solution to the following minimization problem:

$$\min_\theta \parallel \mathbf{y} - \mathbf{U}\beta \parallel + \lambda\beta^T\mathbf{D}\beta \tag{2.62}$$

When $d_k = 0$ and accordingly $\theta_k = 1$, the roughness penalty controlled by λ does not take effect.

It is shown in Craven and Wahba (1979) or Buja, Hastie and Tibshirani (1989) that for the smoother matrix of a cubic smoothing spline, its first two largest eigenvalues are 1's and the associated eigenvectors are spanned by constant $\mathbf{1}$ and $\mathbf{x} := [x_1, x_2, \cdots, x_n]^T$; the left $n-2$ eigenvalues are positive but strictly less than 1 and associated eigenvectors correspond to nonlinear functions or higher order orthogonal polynomials of predictor $\mathbf{x}$ with increasing degree. As noted earlier that the smoother matrix can be viewed as a linear mapping. That is, $\hat{\mathbf{f}} = \mathbf{S}_\lambda \mathbf{y}$ consists of two parts–a linear component due to the orthogonal projection onto the eigenspace spanned by $\mathbf{1}$ and $\mathbf{x}$ and hence a linear function of $\mathbf{x}$; and nonlinear components of nonlinear functions of $\mathbf{x}$. This can be easily seen by partition $\mathbf{U} := (\mathbf{U}_1 \vdots \mathbf{U}_2)$, where $\mathbf{U}_1$ corresponds to the eigenvectors of eigenvalues 1's. $\mathbf{S}_\lambda$ can be written as:

$$\mathbf{S}_\lambda = \mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{D}_2\mathbf{U}_2^T \tag{2.63}$$

where $\mathbf{D}_2$ is a diagonal matrix of the left eigenvalues.

It is easily seen that the first part $\mathbf{U}_1\mathbf{U}_1^T$ is the hat matrix of Least-Squares onto $(\mathbf{1}, \mathbf{x})$ since the eigenvectors are chosen to be an orthonormal basis. That is the smoother matrix $\mathbf{S}_\lambda$ can be partitioned into two parts–*projection* part and *shrinking* part:

$$\mathbf{S}_\lambda = \mathbf{H} + \tilde{\mathbf{S}}_\lambda \tag{2.64}$$

where $\mathbf{H}$ represents the orthogonal projection onto the space spanned by $[\mathbf{1}, \mathbf{x}]$; and $\tilde{\mathbf{S}}_\lambda := (\mathbf{I} - \mathbf{H})\mathbf{S}_\lambda$. It is easily shown that $\tilde{\mathbf{S}}_\lambda$ corresponds to matrix of a shrinking smoother and hence the name of *shrinking* part.

In fact we have $\mathbf{S}_\lambda\mathbf{H} = \mathbf{H}$ due to the fact that $\mathbf{H}$ is the orthogonal projection onto the space spanned by the eigenvectors corresponding to the two eigenvalues 1 of $\mathbf{S}_\lambda$. Then for $\forall \mathbf{y} \in \mathbf{R}^n$:

$$\tilde{\mathbf{S}}_\lambda\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{S}_\lambda\mathbf{y} = \mathbf{y} - \mathbf{H}\mathbf{y}$$

and $\|\tilde{\mathbf{S}}_\lambda\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}\mathbf{y} < \mathbf{y}^T\mathbf{y} = \|\mathbf{y}\|^2$.

Accordingly $\hat{\mathbf{f}}$ can be partitioned into two components as well:

$$\hat{\mathbf{f}} = \hat{\mathbf{g}} + \hat{\tilde{\mathbf{f}}} \tag{2.65}$$

where $\hat{\mathbf{g}} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x}$ with β_0, β_1 as the coefficients from the linear regression of $\mathbf{y}$ on $[\mathbf{1}, \mathbf{x}]$ and $\hat{\tilde{\mathbf{f}}}$ are nonlinear functions of $\mathbf{x}$.

2.6.3 Additive Models: P-Covariate Case

The penalized least-squares criterion for the single-predictor case can be easily generalized to the p-predictor case by minimizing:

$$(\mathbf{y} - \sum_j \mathbf{f}_j)^T(\mathbf{y} - \sum_j \mathbf{f}_j) + \sum_j \lambda_j \mathbf{f}_j^T\mathbf{K}_j\mathbf{f}_j, \tag{2.66}$$

where λ_j is the smoothing parameter, and $\mathbf{K}_j$ is the cubic smoothing-spline penalty matrix for the j^{th} predictor, which is defined analogously to the penalty matrix $\mathbf{K}$ for the single-predictor case as seen in section 1.2.

Based on penalized least-squares criterion, now we differentiate 2.66 with respect to each of $\mathbf{f}_k$'s, we obtain:

$$\hat{\mathbf{f}}_k = (\mathbf{I} + \lambda_k\mathbf{K}_k)^{-1}(\mathbf{y} - \sum_{j \neq k} \hat{\mathbf{f}}_j) \tag{2.67}$$

As noted before, $(\mathbf{I}+\lambda_k\mathbf{K}_k)^{-1}$ is the smoother matrix for the cubic smoothing spline evaluated at unique values of the k^{th} predictor. As it turns out, the solution to the minimization problem 2.66 is cubic smoothing splines for each of the predictors when given the estimates of smooth functions of other predictors.

The Backfitting Algorithm

For the additive models with p covariates using linear smoothers $\mathbf{S}_1, \mathbf{S}_2, \cdots, \mathbf{S}_p$, which can be written as:

$$E(Y|X) = \alpha + \sum_{j=1}^{p} f_j, \tag{2.68}$$

the *backfitting* algorithm (Friedman and Stuetzle, 1981) can be applied. This iterative algorithm is proceeded as:

- Initialize: $\alpha = \frac{1}{n}\sum_{i=1}^{n} y_i$, $f_1^0 = f_2^0 = \cdots = f_p^0 = 0$ and $m = 0$.
  where $f_k = [f_k(x_{k1}), \cdots, f_k(x_{kn})]'$ is n-dimensional vector for $k = 1, \cdots, p$
- Iterate: $m = m + 1$
  - For $j = 1$ to p do:
    $R_j = \mathbf{y} - \alpha - \sum_{k=1}^{j-1} f_j^m - \sum_{k=j}^{p} f_j^{m-1}$,
    and $f_j^m = \mathbf{S}_j(R_j)$
    where $\mathbf{S}_j$ is the smoother matrix for the j^{th} predictor.
  - Until $RSS = \|\mathbf{y} - \alpha - \sum_{j=1}^{p} f_j^m\|^2$ fails to decrease.

where f_j^m denotes the estimate of f_j at the m^{th} iteration. The purpose to center $\mathbf{Y}$ at the first step (i.e., $\alpha = \frac{1}{n}\sum_{i=1}^{n} Y_i$) is to make $\frac{1}{n}\sum_{i=1}^{n} f_j(x_{ji}) = 0$.

Breiman and Friedman (1985) proved through more general context of the ACE algorithm that the solution $\alpha + \sum_{j=1}^{p} f_j^{\infty}$ is unique. Further, Buja, Hastie and Tibshirani (1989 Theorem 5) showed that for symmetric smoother matrices with eigenvalues in $[0, 1]$ the backfitting algorithm always converges, while the estimates $\hat{f}_1, \cdots, \hat{f}_p$ could be non-unique, a phenomenon usually referred to as "concurvity", the analogue of collinearity in linear regression models.

When cubic smoothing splines are used for all of the predictors, for example, the backfitting solution will not be unique if there exists exact collinearity among predictors. Let $\mathcal{M}_1(\mathbf{S}_k)$ represent the space spanned by the eigenvectors of $\mathbf{S}_k$ with eigenvalue 1's, which correspond to linear functions of the k^{th} predictor or the space spanned by $[\mathbf{1}, \mathbf{x}_k]$, $k = 1, 2, \cdots, p$. The exact concurvity can only occur in the space $\mathcal{M}_1(\mathbf{S}_1) + \mathcal{M}_1(\mathbf{S}_2) + \cdots + \mathcal{M}_1(\mathbf{S}_p)$ and it only exists when the predictors are exactly collinear, which will be dealt with in the standard linear regression manner. On the other hand, the nonlinear components of the estimates $\hat{f}_1, \cdots, \hat{f}_p$ are not affected by the concurvity but unique.

Since most smoother matrices are symmetric and have eigenvalues in $[0, 1]$ and hence have both projection and shrinking parts, in other words, reproduce both linear and nonlinear functions of the predictor, this motivates the idea to divide the original backfitting algorithm into two steps: combining all of the orthogonal projection for all of the predictors into one big projection and using only the non-projection (i.e., shrinking) parts of each smoother in the iterative procedure. This is the basic idea of the modified backfitting algorithm, which is proceeded as:

1. Initialize $\tilde{\mathbf{f}}_1, \tilde{\mathbf{f}}_2, \cdots, \tilde{\mathbf{f}}_p$ and let $\tilde{\mathbf{f}}_+ = \tilde{\mathbf{f}}_1 + \tilde{\mathbf{f}}_2 + \cdots + \tilde{\mathbf{f}}_p$.

2. Regress $\mathbf{y} - \tilde{\mathbf{f}}_+$ onto the space $\mathcal{M}_1(\mathbf{S}_1) + \mathcal{M}_1(\mathbf{S}_2) + \cdots + \mathcal{M}_1(\mathbf{S}_p)$.
   Let $\mathbf{H}$ denote the orthogonal projection onto $\mathcal{M}_1(\mathbf{S}_1) + \mathcal{M}_1(\mathbf{S}_2) + \cdots + \mathcal{M}_1(\mathbf{S}_p)$ corresponding to the least-squares regression onto $[\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p]$. This step is just to set $\mathbf{g} = \mathbf{H}(\mathbf{y} - \tilde{\mathbf{f}}_+)$

3. Apply the original backfitting algorithm to $\mathbf{y} - \mathbf{g}$ using the shrinking parts of the smoother matrices (i.e., $\mathbf{S}_j$'s); this step is to yield an updated additive fit: $\tilde{\mathbf{f}}_+ = \tilde{\mathbf{f}}_1 + \tilde{\mathbf{f}}_2 + \cdots + \tilde{\mathbf{f}}_p$.

4. Repeat steps 2 and 3 until $RSS = \|\mathbf{y} - \mathbf{g} - \tilde{\mathbf{f}}_+\|^2$ fails to decrease with the final estimate for the overall fit $\hat{\mathbf{f}} = \mathbf{g} + \tilde{\mathbf{f}}_+$.

In other words, the estimates $\hat{\mathbf{f}}$ can be written as:

$$\begin{aligned} \hat{\mathbf{f}} &= \mathbf{g} + \tilde{\mathbf{f}}_+ \\ &= \alpha + \sum_{j=1}^{p} \hat{\beta}_j x_j + \sum_{j=1}^{p} \tilde{\mathbf{f}}_j \end{aligned} \tag{2.69}$$

where the nonlinear functions $\tilde{\mathbf{f}}_j$ $j = 1, ..., p$ are unique and hence estimable while concurvity (collinearity) can only affect $\sum_{j=1}^{p} \hat{\beta}_j x_j$ and will be dealt with in the same manner as in the standard linear regression using least-squares.

Weighted Penalized Least Squares

The weighted penalized residual sum of squares has the following form:

$$(\mathbf{y} - \sum_j \mathbf{f}_j)^T \mathbf{W} (\mathbf{y} - \sum_j \mathbf{f}_j) + \sum_j \lambda_j \mathbf{f}_j^T \mathbf{K}_j \mathbf{f}_j, \tag{2.70}$$

where $\mathbf{W}$ is a diagonal matrix of weights; λ_j is the smoothing parameter and $\mathbf{K}_j$ is the cubic smoothing-spline penalty matrix for the j^{th} predictor. Direct differentiation of 3.6 with respect to f_j gives the weighted smoothing splines as $\mathbf{S}_j = (\mathbf{I} + \lambda_j \mathbf{W}^{-1} \mathbf{K}_j)^{-1}$, which are no longer symmetric smoothers.

However, the weighted case can be easily mapped back to the weighted case using the transformations: $\mathbf{y}' = \mathbf{W}^{1/2}\mathbf{y}$; $\mathbf{f}_j' = \mathbf{W}^{1/2}\mathbf{f}_j$; $\mathbf{K}_j' = \mathbf{W}^{-1/2}\mathbf{K}_j\mathbf{W}^{-1/2}$, the above weighted penalized residual sum of squares can be expressed as a unpenalized form:

$$(\mathbf{y}' - \sum_j \mathbf{f}_j')^T (\mathbf{y}' - \sum_j \mathbf{f}_j') + \sum_j \lambda_j \mathbf{f}_j'^T \mathbf{K}_j' \mathbf{f}_j', \tag{2.71}$$

The modified backfitting algorithm will then be applied to the transformed data. Accordingly the smoother matrix of the cubic smoothing splines for the j^{th} predictor would be $\mathbf{S}_j' = (\mathbf{I} + \lambda \mathbf{K}_j')$. It is still symmetric with eigenvalues in $[0, 1]$, and the space spanned by the eigenvectors for the two eigenvalues 1 corresponding to linear functions of the j^{th} predictor (Buja, Hastie and Tibshirani, 1989), which means that it has the effect of mapping the partial residual of the form $(\mathbf{y}' - \sum_j \mathbf{f}_j')$ onto the eigenspace spanned by $[\mathbf{1}, \mathbf{x}_j]$ and the eigenspace corresponding to nonlinear functions of $\mathbf{x}_j$.

2.6.4 Generalized Additive Models

Local Scoring Algorithm

The local scoring for generalized additive models with exponential family consists of the following steps:

- Initialize m=0: $\alpha = \mathbf{s}_1 = \mathbf{s}_2 = \cdots = \mathbf{s}_p = 0$

  where $\mathbf{s}_k = [s_k(x_{k1}), \cdots, s_k(x_{kn})]'$ for $k = 1, \cdots, p$.

- Iterate m=m+1 *(Outer Loop)*

  1. Form the adjusted dependent variable:

     $z_i = \eta_i^{m-1} + (y_i - \mu_i^{m-1})(\partial\eta_i^{m-1}/\partial\mu_i^{m-1})$, for $j = 1, 2, \cdots, n$

     where $\eta_i^{m-1} = \alpha^{m-1} + \sum_{k=1}^{3} s_k^{m-1}$

     and $\mu_i^{m-1} = \exp(\eta_i^{m-1})$.

  2. Form the weights: $w_i = (\frac{\partial\mu_i^{m-1}}{\partial\eta_i^{m-1}})^2 (v_i^{m-1})^{-1}$ where v_i^{m-1} is the variance of y_i at the $(m-1)^{th}$ iteration.

  3. Fit an weighted additive model to z_i using the backfitting algorithm to get the estimates of α^m and s_k^m. *(Inner Loop)*

     (a) For $k = 1, \cdots, p$:

     $$\hat{s}_k^m = \mathbf{S}_k'(z_k - \alpha^{m-1} - \sum_{j \neq k} s_j^{m-1})$$

     where $\mathbf{S}_k' = (I + \lambda_k \mathbf{K}_k')^{-1}$ with $\mathbf{K}_k' = \mathbf{W}^{-1/2}\mathbf{K}_k\mathbf{W}^{-1/2}$ as the penalty matrices for the k^{th} predictor defined in the same manner as deriving the cubic smoothing spline for a single predictor appearing in 2.56; and $\mathbf{W} = \text{diag}\{w_1, w_2, \cdots, w_n\}$ as the weight matrix.

     (b) Compute the backfitting convergence criterion in the inner loop:

     $$\text{RSS}^m = \frac{1}{n}\sum_i (z_i - \alpha^m - \sum_k \hat{s}_k^m)^2$$

     (c) Stop when RSS^m fails to decrease.

     (d) Compute $\eta_i^m = \alpha^m + \sum_k \hat{s}_k^m$ and $\mu_i^m = \exp(\eta_i^m)$.

- Stop when the deviance $\text{dev}(Y, \mu^m)$ fails to decrease.

The inner loop of backfitting algorithm within the outer loop of local scoring algorithm is just repeatedly smoothing the adjusted dependent variable on a single coordinate using the weighted penalized least-squares criterion.

Chapter 3

Smooth Age-Period-Cohort Models Using Smoothing Splines

3.1 Holford's Parameterization to the Smooth APC Models

Holford's parameterization was originally proposed to the data tabulated in five-year age and period intervals, but many authors have suggested the use of this parameterization to the yearly data. In this section, we review how Holford's parameterization (1980) can be applied to the yearly data. We use $i = 1, ..., n_a$, $j = 1, ..., n_p$ and $k = j - i + n_a$ ($\equiv 1, ..., n_c = n_a + n_p - 1$) to index age, period and cohort respectively. Let a_i and p_j be the actual yearly values in the two way table by age and period, which is also called the Lexis diagram; the associated birth cohort c_k will have a length of two years, which will be identified by the mid-point of the intervals through $c_k = p_j - a_i$. For example, the birth cohort corresponding to those who were 50 years old in 1980 and so were born in $1929 - 1931$ will be identified by 1930.

Assuming the count in each cell of the Lexis diagram $y_{ij} \sim \text{Poisson}(\mu_{ij})$, the expected rate $\lambda_{ijk} = \mu_{ij}/N_{ij}$, where N_{ij} is the person-years at risk. $\log(N_{ij})$ would be treated as offset when fitting a log-linear poisson regression model with $\hat{\lambda}_{ijk} = y_{ij}/N_{ij}$ as the response. The general form of the smooth APC model is:

$$\log(\lambda_{ijk}) = \mu + f_a(a_i) + f_p(p_i) + f_c(c_k), \tag{3.1}$$

where μ represents the intercept, $f_a()$, $f_p(.)$ and $f_c(.)$ are arbitrary smooth functions representing the effect of age, period and cohort respectively. The usual constraints apply here: $\sum_i f_a(a_i) = \sum_j f_p(p_j) = \sum_k f_c(c_k) = 0$. Holford's parameterization is to partition each effect into two components–linear trend and curvature effect with two trends being orthogonal

to each other. For example, we can represent the effect of age $f_a(a_i)$ as:

$$f_a(a_i) = \bar{a}_i\beta_a + \tilde{f}_a(a_i) \tag{3.2}$$

where $\bar{a}_i = a_i - \sum_i a_i/n_a$ is the centralized values for age, β_a is the slope of the linear trend of age and $\tilde{f}_a(a_i)$ is the curvature effect. Let $\mathbf{M}_a$ be the basis matrix for orthogonal polynomials or restricted regression splines (i.e., natural cubic spline) evaluated at $\mathbf{a} = (a_1, ..., a_{n_a})'$. The matrix corresponding to the curvature effects can be obtained by projecting the columns of $\mathbf{M}_a$ onto the orthogonal complement of $[\mathbf{1}|\mathbf{a}]$, i.e., the linear trend being removed from $\mathbf{M}_a$. The "detrend()" function of R in the "Epi" package by Carstensen (2007) implements this projection. The resulting matrix $\tilde{\mathbf{M}}_a$ has two fewer columns than $\mathbf{M}_a$ with the corresponding coefficient vector $\tilde{\beta}_a$. Let $\tilde{\mathbf{M}}_a[i,]$ represent the i^{th} row of $\tilde{\mathbf{M}}_a$, then the function representing curvature effects of age can be expressed as: $\tilde{f}_a(a_i) = \tilde{\mathbf{M}}_a[i,]\tilde{\beta}_a$. Similarly we obtain the "detrended" matrices $\tilde{\mathbf{M}}_p$ and $\tilde{\mathbf{M}}_c$ and define $\tilde{f}_p(p_j) = \tilde{\mathbf{M}}_p[j,]\tilde{\beta}_p$ and $\tilde{f}_c(c_k) = \tilde{\mathbf{M}}_c[k,]\tilde{\beta}_c$ for period and cohort respectively. We can rewrite the equation (3.3) as

$$\begin{aligned}
\log(\lambda_{ijk}) &= \mu + \bar{a}_i\beta_a + \bar{p}_j\beta_p + \bar{c}_k\beta_c + \tilde{f}_a(a_i) + \tilde{f}_p(p_j) + \tilde{f}_c(c_k), \\
&= \mu + \bar{a}_i(\beta_a - \beta_c) + \bar{p}_j(\beta_p + \beta_c) + \tilde{f}_a(a_i) + \tilde{f}_p(p_j) + \tilde{f}_c(c_k) \qquad (3.3) \\
\text{or} &= \mu + \bar{a}_i(\beta_a + \beta_c) + \bar{c}_k(\beta_p + \beta_c) + \tilde{f}_a(a_i) + \tilde{f}_p(p_j) + \tilde{f}_c(c_k)
\end{aligned}$$

because $\bar{c}_k = \bar{p}_j - \bar{a}_i$. The slope of the linear effect representing the overall increasing or decreasing trend for each temporal factor are not estimable. However, certain linear functions of these slopes such as $\beta_a - \beta_c$, $\beta_a + \beta_c$ and $\beta_p + \beta_c$ are all estimable. The curvature effects representing the deviations from the overall trend for each temporal factor are always estimable, i.e., the three coefficient vectors $\tilde{\beta}_a$, $\tilde{\beta}_p$ and $\tilde{\beta}_c$ are all estimable.

When using cubic smoothing splines, we can still apply Holford's parameterization except the functions of curvature effects $\tilde{\mathbf{f}}_a$, $\tilde{\mathbf{f}}_p$ and $\tilde{\mathbf{f}}_c$ are smooth functions that do not have parametric representations.

3.2 Smooth APC Models Using Smoothing Splines

In this section we review some properties of the smoothing splines that are helpful in explaining the parameterization of the APC model when using smoothing splines. Specifically, cubic smoothing splines will be focused here.

3.2.1 smoothing splines

A cubic smoothing spline is the minimizer of the penalized least-squares criterion:

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt, \tag{3.4}$$

over all functions with two continuous derivatives. λ known as the smoothing parameter trades off the fit as well as the smoothness: when $\lambda = 0$ the solution would be any function that interpolate the points; when $\lambda = \infty$, the solution is a least-squares line. For any $\lambda \in (0, \infty)$ the solution is a natural cubic spline with knots at each distinct values for $\mathbf{x}$. Green and Silverman give the complete and detailed proof (1994) to this statement. Using this conclusion one can show that when given the value of smoothing parameter λ, the cubic smoothing spline is a linear smoother in the sense that the fitted value can be written as $\hat{\mathbf{f}} = (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{y} = \mathbf{S}_\lambda\mathbf{y}$, where $\mathbf{K}$ is the penalty matrix and $\mathbf{S}_\lambda$ is called the smoother matrix. Both matrices depend only on the observed values for $\mathbf{x}$ but not on $\mathbf{y}$. This presentation indicates the minimization problem (3.4) can be reduced to:

$$(\mathbf{y} - \mathbf{f})^T(\mathbf{y} - \mathbf{f}) + \lambda\mathbf{f}^T\mathbf{K}\mathbf{f}. \tag{3.5}$$

Hastie and Tibshirani (1990, §2.10) give the details.

It can be also shown that the smoother matrix $\mathbf{S}_\lambda$ is symmetric and have eigenvalues in $[0, 1]$: the first two are 1's with the corresponding eigen-space being spanned by $[\mathbf{1}, \mathbf{x}]$; the left are strictly less than 1 but greater than 0. This fact leads to the eigenvalue-decomposition of $\mathbf{S}$: $\mathbf{S}_\lambda = \mathbf{U}_1\mathbf{U}_1^T + \mathbf{U}_2\mathbf{D}_2\mathbf{U}_2^T = \mathbf{H} + \tilde{\mathbf{S}}_\lambda$, where $\mathbf{U}_1\mathbf{U}_1^T$ is the least-squares hat matrix when the eigenvectors in $\mathbf{U}_1$ and $\mathbf{U}_2$ are orthonormal basis. This decomposition suggests that the fitted value $\hat{\mathbf{f}}$ has two components corresponding to linear and nonlinear functions of $\mathbf{x}$:

$$\hat{\mathbf{f}} = \hat{\mathbf{g}} + \hat{\tilde{\mathbf{f}}}$$

where, $\hat{\mathbf{g}} = \hat{\beta}_0\mathbf{1} + \hat{\beta}_1\mathbf{x}$ with coefficients β_0, β_1 from the ordinary least-squares regression of $\mathbf{y} - \hat{\tilde{\mathbf{f}}}$ on $[\mathbf{1}, \mathbf{x}]$; the nonlinear component $\hat{\tilde{\mathbf{f}}}$ is the nonlinear function of $\mathbf{x}$. The estimation of the two components are accomplished in the orthogonal spaces, so we have $\mathbf{H}\hat{\mathbf{f}} = \mathbf{0}$ and $\tilde{\mathbf{S}}_\lambda\hat{\mathbf{g}} = \mathbf{0}$.

When the cubic smoothing splines are applied to the model assuming the response have a exponential family density, i.e., the mean $\mu = \mathrm{E}(Y|X_1, ..., X_p)$ being linked to the predictors

through $g(\mu) = \beta_0 + \sum_j f_j(X_j)$, where $f_j(X_j)$ will be estimated by smoothing splines. Such models are often called Generalized Additive Models (GAMs). The estimation of β_0 and f_1,...,f_p can be done by the local scoring with the backfitting algorithm as the inner loop. The local scoring algorithm generates the working response variables and the weights in each outer iteration and the backfitting algorithm in the inner loop is to fit a weighted additive models on the generated response variable. More specifically, the backfitting algorithm for GAM with smoothing splines is to find the minimizer to the following form of weighted penalized least-squares criterion:

$$(\mathbf{z} - \sum_j \mathbf{f}_j)^T \mathbf{W} (\mathbf{z} - \sum_j \mathbf{f}_j) + \sum_j \lambda_j \mathbf{f}_j^T \mathbf{K}_j \mathbf{f}_j, \tag{3.6}$$

where $\mathbf{z}$ is the working response variable; $\mathbf{W}$ is a diagonal matrix of the weights; λ_j and $\mathbf{K}_j$ are the smoothing parameter and the penalty matrix for the j^{th} predictor. Direct differentiation of 3.6 with respect to f_j gives the weighted smoothing splines as $\mathbf{S}_j = (\mathbf{I} + \lambda_j \mathbf{W}^{-1} \mathbf{K}_j)^{-1}$. When using the transformations $\mathbf{y}' = \mathbf{W}^{1/2}\mathbf{y}$; $\mathbf{f}_j' = \mathbf{W}^{1/2}\mathbf{f}_j$; $\mathbf{K}_j' = \mathbf{W}^{-1/2}\mathbf{K}_j\mathbf{W}^{-1/2}$, the criterion (3.6) can be reduced to an unweighted form with the smoother matrix for the j^{th} predictor as $\mathbf{S}_j' = (\mathbf{I} + \lambda \mathbf{K}_j')^{-1} (= \mathbf{W}^{1/2}\mathbf{S}_j\mathbf{W}^{-1/2})$. Having this relationship, it can be easily seen that $\mathbf{S}_j'$ is symmetric with eigenvalues in $[0, 1]$, and the two eigenvalues 1's correspond to linear functions of $\mathbf{W}^{1/2}[\mathbf{1}, \mathbf{x}_j]$. Hence the overall estimates in the transformed coordinates can be also separated into two parts–corresponding to linear and nonlinear functions of $\mathbf{W}^{1/2}[\mathbf{1}, \mathbf{x}_1, ..., \mathbf{x}_p]$:

$$\mathbf{W}^{1/2}\hat{\mathbf{f}} = \mathbf{W}^{1/2}\mathbf{1}\hat{\beta}_0 + \sum \mathbf{W}^{1/2}\mathbf{x}_j\hat{\beta}_j + \sum \hat{\tilde{\mathbf{f}}}_j \tag{3.7}$$

where $\hat{\beta}_0, \hat{\beta}_1$,...,$\hat{\beta}_p$ are the coefficients from the least-squares regression of $\mathbf{z} - \sum \hat{\tilde{\mathbf{f}}}_j$ on $\mathbf{W}^{1/2}[\mathbf{1}, \mathbf{x}_1, ..., \mathbf{x}_p]$. As in the single-predictor case, the two parts are estimated in the orthogonal spaces: i.e., $\hat{\tilde{\mathbf{f}}}_j' \hat{\beta}_j \mathbf{W}^{1/2}\mathbf{x}_j = 0$. When transformed back to the original coordinate, (3.7) can be written as:

$$\begin{aligned} \hat{\mathbf{f}} &= \mathbf{1}\beta_0 + \sum \mathbf{x}_j\beta_j + \sum \mathbf{W}^{-1/2}\tilde{\mathbf{f}}_j \\ &= \mathbf{1}\beta_0 + \sum \mathbf{g}_j + \sum \tilde{\mathbf{f}}_j^* \end{aligned} \tag{3.8}$$

where $\mathbf{g}_j = \mathbf{x}_j\beta_j$ and $\tilde{\mathbf{f}}_j^* = \mathbf{W}^{-1/2}\tilde{\mathbf{f}}$, which are referred to as the linear and nonparametric components respectively in gam package in R. Based on the conclusion concerning the convergence of the backfitting algorithm for Generalized Additive Models by Buja, Hastie and Tibshirani (1989), in case of any linear dependency among the p predictors, only the

estimation of β_1,...,β_p will be affected and handled in the same fashion as in ordinary least-squares regression with $\tilde{\mathbf{f}}_j$ ($j = 1, ..., p$) being all estimable. The orthogonal relationship between the two components will be:

$$\tilde{\mathbf{f}}_j^{*'}\mathbf{W}\mathbf{g}_j = \mathbf{0}, \tag{3.9}$$

indicating that additional linear components can be further extracted from the nonparametric components.

3.2.2 The Parameterization Using Smoothing Splines

When cubic smoothing splines are applied on all three predictor: Age, Period and Cohort, the contribution of each predictor can be decomposed into two components–a linear component (a weighted least square fit) and a nonparametric component:

$$\begin{aligned}
\hat{f}_a(a_i) &= \delta_a \bar{a}_i + \tilde{f}_a(a_i), \\
\hat{f}_p(p_j) &= \delta_p \bar{p}_j + \tilde{f}_p(p_j), \\
\hat{f}_c(c_k) &= \delta_c \bar{c}_k + \tilde{f}_c(c_k),
\end{aligned} \tag{3.10}$$

where $\bar{a}_i$, $\bar{p}_j$ and $\bar{c}_k$ are the centralized values for age, period and cohort respectively; δ_a, δ_p and δ_c are the slopes from the weighted least-squares regression, in which the linear dependency $\bar{c}_k = \bar{p}_j - \bar{a}_i$ will make the design matrix in this step singular and how "gam" function in gam package of R deals with this is to force one of the slopes to be zero depending on the order in which the three variables are specified. For example, when cohort is the last one to be specified in the model, the estimate of the cohort slope $\hat{\delta}_c = 0$, i.e., $\delta_a - \delta_c$, $\delta_p + \delta_c$ and $\delta_a + \delta_p$ are estimable but δ_a, δ_p or δ_c are not. $\tilde{f}_a$, $\tilde{f}_p$ and $\tilde{f}_c$ are the estimable nonparametric components. Note the expression (3.9), we can extract additional linear components from each of them:

$$\begin{aligned}
\tilde{f}_a(a_i) &= \delta_a^* \bar{a}_i + f_a^{cur}(a_i), \\
\tilde{f}_p(p_i) &= \delta_p^* \bar{p}_j + f_p^{cur}(p_j), \\
\tilde{f}_c(c_i) &= \delta_c^* \bar{c}_k + f_c^{cur}(c_k),
\end{aligned} \tag{3.11}$$

where the three slopes δ_a^*, δ_p^* and δ_c^* as well as all three curvature components $f_a^{cur}(a_i)$, $f_p^{cur}(p_i)$ and $f_c^{cur}(c_i)$ are uniquely determined and have liner components removed. Substituting the

above equations (3.11) into (3.10), we will get:

$$\begin{aligned}\log(\lambda_{ijk}) = \mu &+ (\beta_a - \beta_c)\bar{a}_i + f_a^{cur}(a_i) \\ &+ (\beta_p + \beta_c)\bar{p}_j + f_p^{cur}(p_j) \\ &+ f_c^{cur}(c_k)\end{aligned} \tag{3.12}$$

or

$$\begin{aligned}\log(\lambda_{ijk}) = \mu &+ (\beta_a + \beta_p)\bar{a}_i + f_a^{cur}(a_i) \\ &+ f_p^{cur}\bar{p}_j \\ &+ (\beta_p + \beta_c)\bar{c}_k + f_c^{cur}(c_k)\end{aligned} \tag{3.13}$$

where

$$\begin{aligned}\beta_a &= \delta_a + \delta_a^* \\ \beta_p &= \delta_p + \delta_p^* \\ \beta_c &= \delta_c + \delta_c^*\end{aligned} \tag{3.14}$$

with β_a, β_p and β_c are the true slopes (i.e., representing the linear trends) of age, period and cohort respectively. Based on the above analysis, it is easily seen that the linear functions $\beta_a - \beta_c$, $\beta_p + \beta_c$ and $\beta_p + \beta_c$ along with f_a^{cur}, f_p^{cur} and f_c^{cur} are all estimable but individually β_a,β_p and β_c are not. In conclusion, the estimable functions proved by Holford (1983) for the traditional APC model still hold here when GAM was fitted using smoothing splines.

Chapter 4

Simulation Study and Data Analysis

4.1 Simulation Study

Simulations with pre-specified age, period and cohort effects were conducted to compare the estimations from three different modeling strategies–the new APC model using smoothing spline (SS), the APC model using natural spline (NS) by Heuer (1997) and the APC model using orthogonal polynomials (OP), which is basically using the same idea as Holford's (1983) parametrization on the traditional APC factor model but will be applied here to the yearly data. The three methods will be referred to as GAM+SS, GLM+NS and GLM+OP for convenience, in which the first acronym (GLM or GAM) represents the framework of model fitting while the second acronym represents the smoothing techniques.

The age curve is given by the following function:

$$f_a(a_i) = \beta_a(a_i - \bar{a}),\ a_i = 51, ..., 80. \tag{4.1}$$

where $\bar{a} = \sum_i a_i$ and β_a is the true linear trend of age effects. The values chosen for β_a are -0.05, 0 and 0.05. The assumed scenarios for age effects are reflections of the pattern in the age distribution for most cancers; that is, the risk is linearly increasing ($\beta_a > 0$) or decreasing ($\beta_a < 0$) for older people (in our case between 51 and 80 years old) while in some other cases, the risk of getting the cancer or not could have nothing to do with the age ($\beta_a = 0$).

The period curve is given by

$$f_p(p_j) = \beta_p(p_j - \bar{p}), \tag{4.2}$$

for periods p_j starting 1985 up to year 2004, where $\bar{p} = \sum_j p_j$ and β_p is the true linear trend of period effects. The values chosen for β_p are -0.05, 0 and 0.05. The assumed scenarios

for period effects could also be realistic. When the effective treatments are introduced the incidence rate could be decreasing ($\beta_p < 0$) while there would be periods when the incidence rate remains fairly stable ($\beta_p = 0$) or increasing ($\beta_p > 0$).

Three different patterns for the birth cohort effects are assumed, starting the birth year of 1905 up to 1953. The cohort curve I:

$$f_c(c_k) = 1.25 + 0.15(\frac{c_k - \bar{c}}{15}) + 0.15(\frac{c_k - \bar{c}}{15})^2 + 0.025(\frac{c_k - \bar{c}}{15})^4, \tag{4.3}$$

the cohort curve II:

$$f_c(c_k) = 1.25 + 0.15(\frac{c_k - \bar{c}}{15}) - 0.15(\frac{c_k - \bar{c}}{15})^2 - 0.025(\frac{c_k - \bar{c}}{15})^4, \tag{4.4}$$

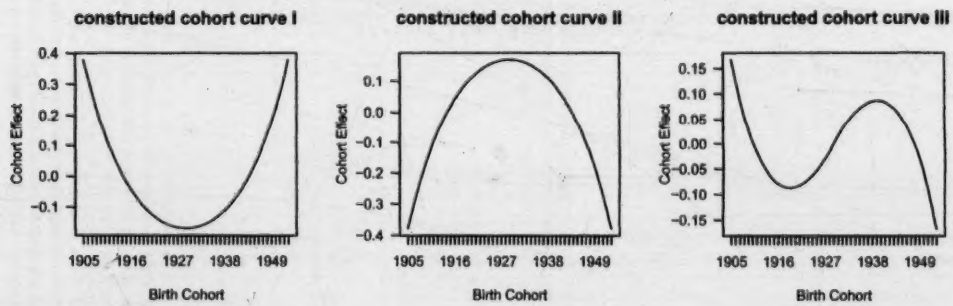
and the cohort curve III:

$$f_c(c_k) = \begin{cases} 1.25 + 0.15(\frac{c_k - \bar{c}}{15}) - 0.15(\frac{c_k - \bar{c}}{15})^2 - 0.025(\frac{c_k - \bar{c}}{15})^4, \\ 1.25 + 0.15(\frac{c_k - \bar{c}}{15}) + 0.15(\frac{c_k - \bar{c}}{15})^2 + 0.025(\frac{c_k - \bar{c}}{15})^4, \end{cases} \tag{4.5}$$

where $c_k = p_j - a_i$ and $\bar{c} = \sum_k c_k$.

Birth cohort effects can be thought of as the exposure to some underlying but often unknown risk factors, and hence the pattern of the birth cohort curve could be somewhat complicated. The assumed cohort curve I, II and III are shown in 4.1. The assumed cohort curve I is a reflection of the situation when the incidence rate for earlier cohorts born between 1905 and 1929 is decreasing and yet for the more recent cohorts born between 1929 and 1953 there was an increasing risk; the assumed cohort curve II is a reflection of the opposite situation when there was an increase in the incidence rates for earlier cohorts and yet an decrease for the more recent cohorts; the assumed cohort curve III is a periodic extension of cohort curve I and II.

Figure 4.1: Constructed Birth Cohort Effect Curves I, II and III.



For each of the chosen cohort curves above, the simulation is done for the three different values representing the linear trends for age and period effects. The values of coefficients in all assumed curves for age, period and cohort are chosen to simulate the situation when we would expect very low incidence and approximately 30% of zero incidences in the lexis diagram.

The expected frequencies μ_{ij} per $100,000$ in the two-way table by age and period are given by:

$$\mu_{ij} = \exp[\log(N_{ij}) + f_a(a_i) + f_p(p_j) + f_c(c_k)] \tag{4.6}$$

where $i = 1, ..., 30$, $j = 1, ..., 30$ and $k = 1, ..., 59$ are the indices for age, period and birth cohort respectively; a_i and p_j are the age and period in years with c_k determined by $p_j - a_i$, the mid-point of the corresponding birth cohort in two-year intervals. Here, we set $N_{ij} = 0.2$ for all i and j, i.e., the population figure is assumed to be $20,000$ for all cells of the two-way table and we re-scaled it to 0.2. Using the constructed cell frequencies, we sampled the data 1000 times from these Poisson distributions, calculating the 95% empirical confidence intervals for the estimable cross-sectional age slope (i.e., $\beta_a - \beta_c$), overall slope of time trend (i.e., $\beta_a - \beta_c$) as well as the curvature effects for birth cohort by selecting the 2.5 and 97.5 percentiles of the corresponding 1000 estimates. The estimable age and period slopes in the two way table by Age (i.e., Lexis diagram) can be obtained by forcing the linear trend in the birth cohort effect to be zero, and hence the cross-sectional age slope and overall slope of time trend represent the time trend in both period and cohort effect.

Holford's parametrization proposed to the traditional APC model treats the three factors age, period and cohort as categorical variables, similar to a factorial design. The data are usually aggregated into five-year age and period intervals, but this idea can be directly applied to the yearly ungrouped data. Holford's parametrization uses orthogonal polynomials to partition the effects into two broad components–the linear effects and curvature effects. While the three linear slopes representing the corresponding linear effects are not individually estimable, their curvature components are uniquely determined, which reflect a sudden change in the overall trend and might be related to some important changes in other aspects. For the estimable curvature components, they can be further divided into quadratic, cubic and even higher-order components. Since each of these components is also estimable, one can choose to include only

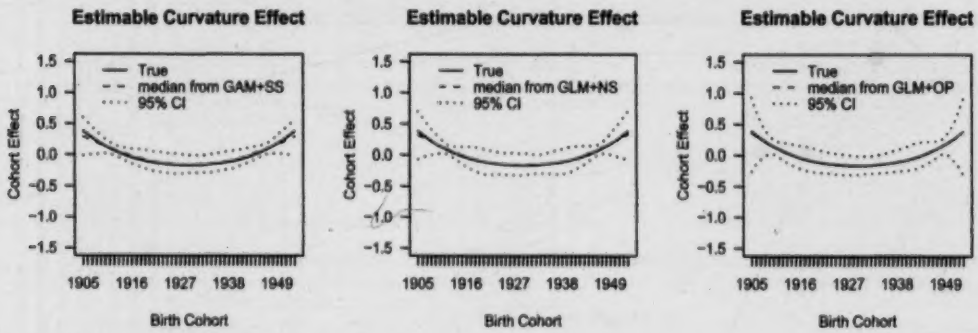
some components from the overall curvature effects into the model. Since the highest order of the assumed curves for birth cohort is 4, we chose to use 4 degrees of freedom for birth cohort, including up to the 4^{th} order of orthogonal polynomials for birth cohort, without constant term. For comparison we also chose to use 4 degrees of freedom on the birth cohort for GAM+SS and GLM+NS. The smoothing parameter λ_c will be chosen to make $df_c = \text{trace}[\mathbf{S}_c(\lambda_c)] - 1 \equiv 4$, where $\mathbf{S}_c$ is the smoothing spline matrix constructed for birth cohort in GAM+SS. Three inner knots at equally spaced quantiles of the values of birth cohort (i.e., at 25%, 50% and 75% percentiles) will be chosen to construct the basis matrix without including intercept for the natural cubic spline when using GLM+NS.

The mean square errors (MSEs) of the estimates of cross-sectional age slope and overall slope of time trend reported in Table 4.1 through 4.27will be used to access the model performance. The medians of estimated curvature effects and their 95% confidence intervals are shown in figure 4.2 through 4.28. All of the medians of the curves are close enough to the true curve and according to the criteria of MSEs the estimates of two slopes are also good enough for each method. However, GAM+SS always returns lower MSEs and narrower and more stable empirical 95% confidence intervals compared to GLM+NS or GLM+OP. By comparing the empirical variances and MSEs, the bias from GAM+SS is smaller than the bias from GLM+NS or GLM+OP. Therefore, in the situations of low incidence and even many zero incidence, generalized additive APC model using smoothing splines would give more stable estimates and should be recommended although both GLM+NS and GLM+OP are also nice alternatives. Here we are considering the situation when the model to be fitted reflects the true situation; that is, we assume there is only linear trend in both age and period effects in the expected incidence rate and hence we only add linear terms of age and period into the model. In other words, we use almost "correct" models to check the stability of each smoothing technique. We admit that in real life it is impossible for us to know the true situation. The test of robustness of these methods is forthcoming.

Table 4.1: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	4.02	3.2702	3.2749
		GLM+NS	4.03	3.5872	3.5948
		GLM+OP	4.03	3.8006	3.8091
$\beta_p + \beta_c$	6	GAM+SS	6.00	6.7952	6.7953
		GLM+NS	6.01	6.8472	6.8488
		GLM+OP	6.01	6.9676	6.9684

Figure 4.2: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

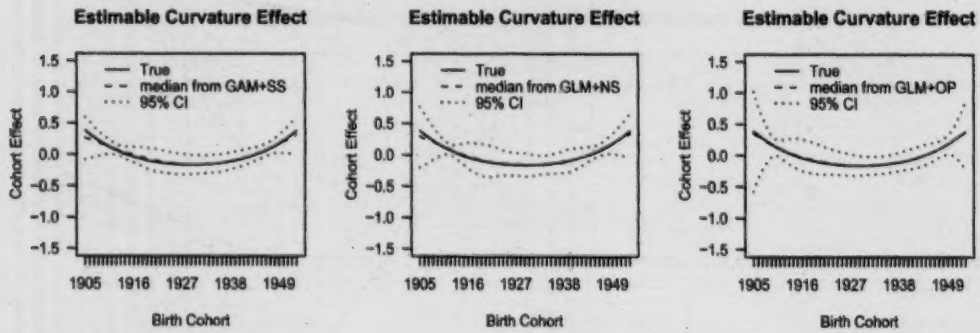


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.2: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	-1.08	3.6956	3.7537
		GLM+NS	-1.05	4.1809	4.2050
		GLM+OP	-1.06	4.5319	4.5694
$\beta_p + \beta_c$	6	GAM+SS	6.04	6.7135	6.7270
		GLM+NS	6.01	6.8407	6.8420
		GLM+OP	6.02	7.0154	7.0215

Figure 4.3: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

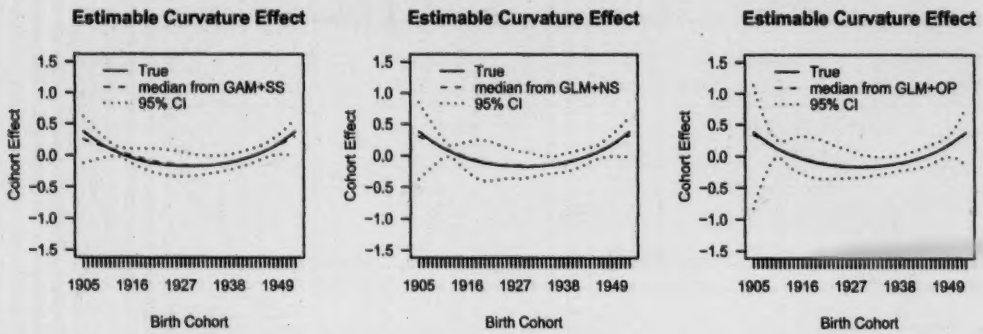


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.3: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-6	GAM+SS	-6.17	4.1103	4.3928
		GLM+NS	-6.14	5.2706	5.4633
		GLM+OP	-6.17	6.2161	6.5053
$\beta_p + \beta_c$	6	GAM+SS	6.14	7.6804	7.8757
		GLM+NS	6.10	8.5383	8.6386
		GLM+OP	6.14	9.4194	9.6052

Figure 4.4: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

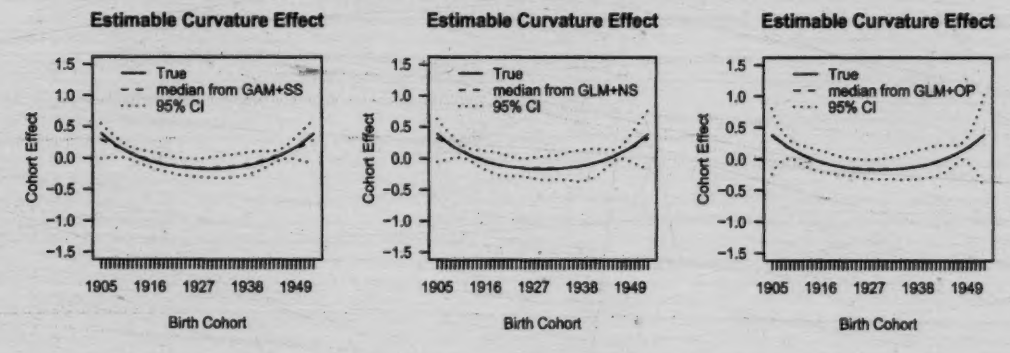


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.4: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	4.09	3.3656	3.4405
		GLM+NS	4.07	3.7775	3.8333
		GLM+OP	4.08	4.0229	4.0881
$\beta_p + \beta_c$	1	GAM+SS	1.00	6.1804	6.1805
		GLM+NS	1.02	6.2072	6.2105
		GLM+OP	1.01	6.3684	6.3692

Figure 4.5: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$.

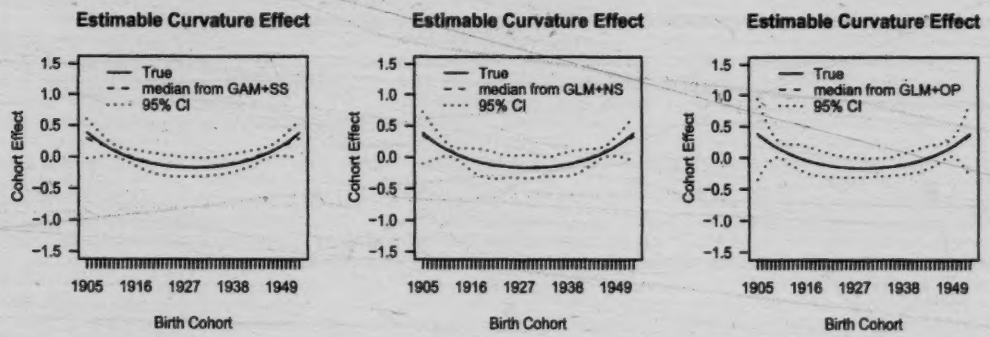


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.5: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	-1.02	3.4011	3.4064
		GLM+NS	-1.02	3.6946	3.6971
		GLM+OP	-1.02	3.9353	3.9396
$\beta_p + \beta_c$	1	GAM+SS	0.98	6.6411	6.6450
		GLM+NS	0.97	6.6164	6.6255
		GLM+OP	0.97	6.7514	6.7579

Figure 4.6: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$.

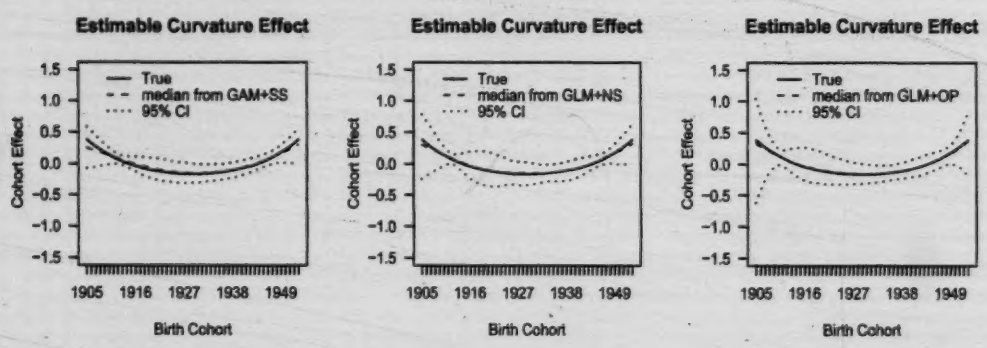


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.6: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-6	GAM+SS	-6.10	3.6887	3.7897
		GLM+NS	-6.07	4.3084	4.3580
		GLM+OP	-6.08	4.7909	4.8520
$\beta_p + \beta_c$	1	GAM+SS	1.06	6.8969	6.9302
		GLM+NS	1.02	7.1710	7.1735
		GLM+OP	1.03	7.5035	7.5124

Figure 4.7: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve I with $\beta_c = 0.01$.

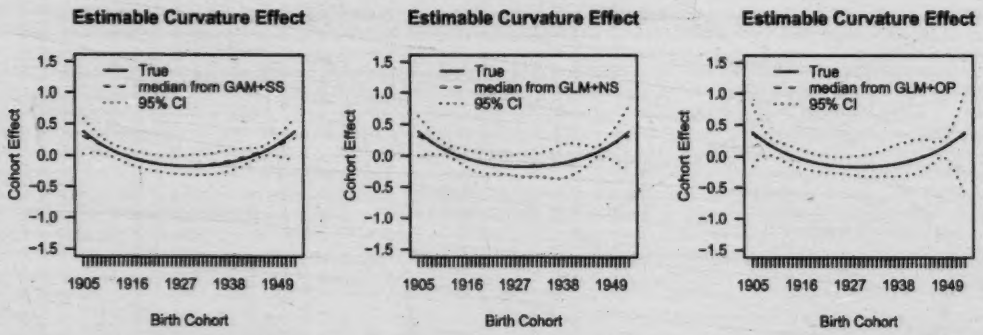


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.7: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	4.06	3.7330	3.7659
		GLM+NS	4.02	4.2495	4.2534
		GLM+OP	4.03	4.5857	4.5936
$\beta_p + \beta_c$	-4	GAM+SS	-4.08	7.3084	7.3687
		GLM+NS	-4.04	7.5435	7.5575
		GLM+OP	-4.05	7.8790	7.9024

Figure 4.8: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

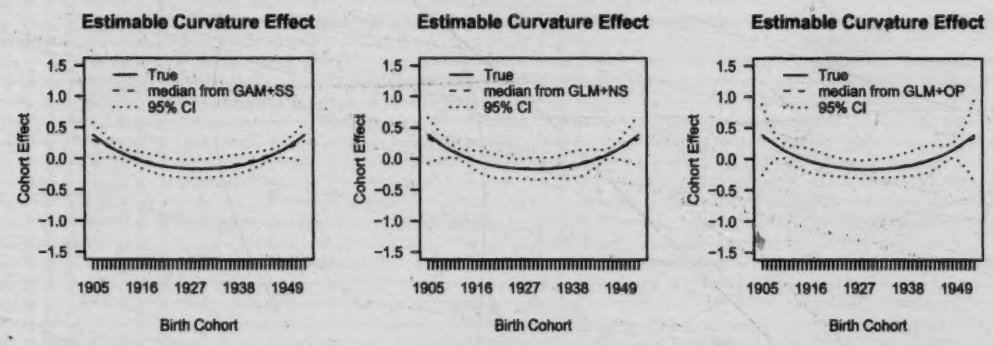


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.8: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	-0.95	3.2213	3.2515
		GLM+NS	-0.96	3.5273	3.5457
		GLM+OP	-0.95	3.7799	3.8033
$\beta_p + \beta_c$	-4	GAM+SS	-4.03	6.0739	6.0818
		GLM+NS	-4.02	6.0165	6.0217
		GLM+OP	-4.03	6.1690	6.1760

Figure 4.9: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

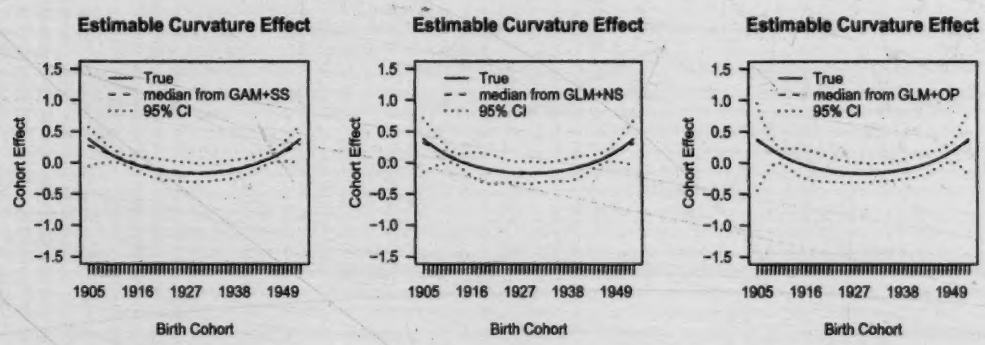


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.9: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-6	GAM+SS	-6.01	3.3651	3.4248
		GLM+NS	-6.00	3.7122	3.7560
		GLM+OP	-6.00	3.9854	4.0335
$\beta_p + \beta_c$	-4	GAM+SS	-3.95	6.9145	6.9217
		GLM+NS	-3.98	6.9512	6.9513
		GLM+OP	-3.97	7.1125	7.1130

Figure 4.10: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve I with $\beta_c = 0.01$.

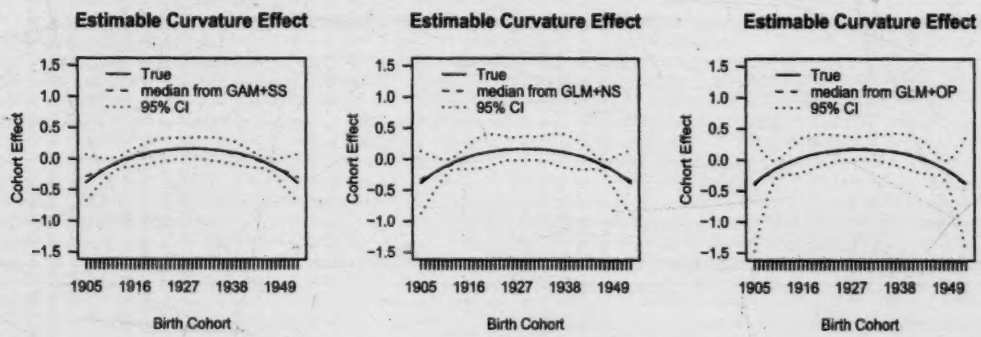


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.10: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value	Model	Empirical Mean of estimates	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	4.03	4.7680	4.7801
		GLM+NS	4.04	6.0457	6.0591
		GLM+OP	4.03	7.0689	7.0768
$\beta_p + \beta_c$	6	GAM+SS	5.98	8.2204	8.2248
		GLM+NS	5.96	8.9264	8.9422
		GLM+OP	5.97	9.7308	9.7374

Figure 4.11: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

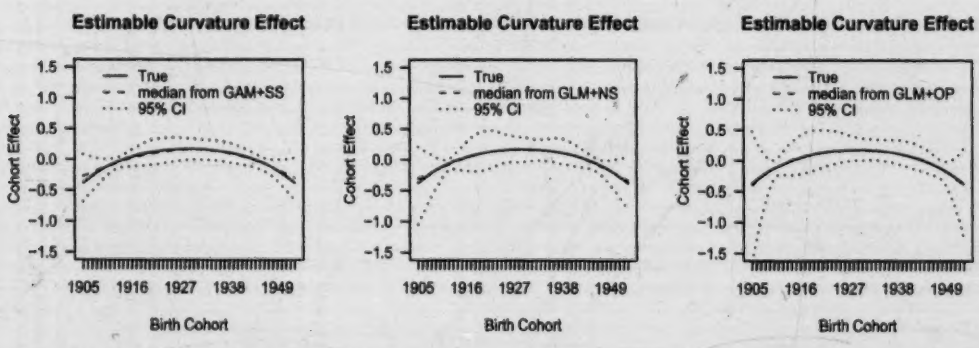


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.11: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	-0.95	5.1810	5.2044
		GLM+NS	-1.03	6.8296	6.8366
		GLM+OP	-1.07	8.1945	8.2402
$\beta_p + \beta_c$	6	GAM+SS	5.98	9.5195	9.0559
		GLM+NS	6.05	10.0149	10.0474
		GLM+OP	6.09	10.6144	11.3423

Figure 4.12: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

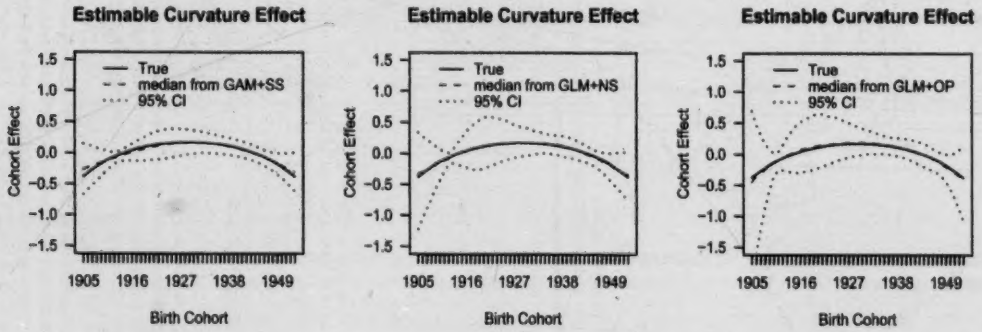


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.12: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-6	GAM+SS	-6.00	5.6359	5.6376
		GLM+NS	-6.17	8.9310	9.2083
		GLM+OP	-6.24	11.5550	12.1506
$\beta_p + \beta_c$	6	GAM+SS	6.00	9.2956	9.2977
		GLM+NS	6.17	12.2819	12.5599
		GLM+OP	6.24	14.7242	15.2888

Figure 4.13: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

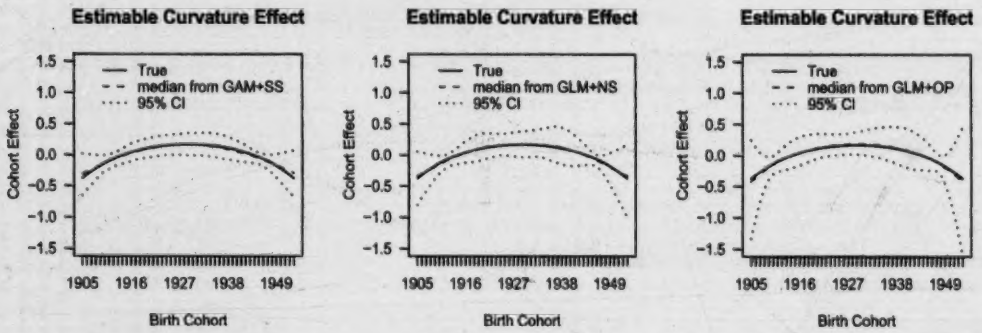


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.13: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	4.01	5.0849	5.0867
		GLM+NS	4.07	6.4702	6.5149
		GLM+OP	4.10	7.6514	7.7469
$\beta_p + \beta_c$	1	GAM+SS	0.99	9.0591	9.0597
		GLM+NS	0.93	9.8598	9.9060
		GLM+OP	0.91	11.0543	11.1398

Figure 4.14: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$.

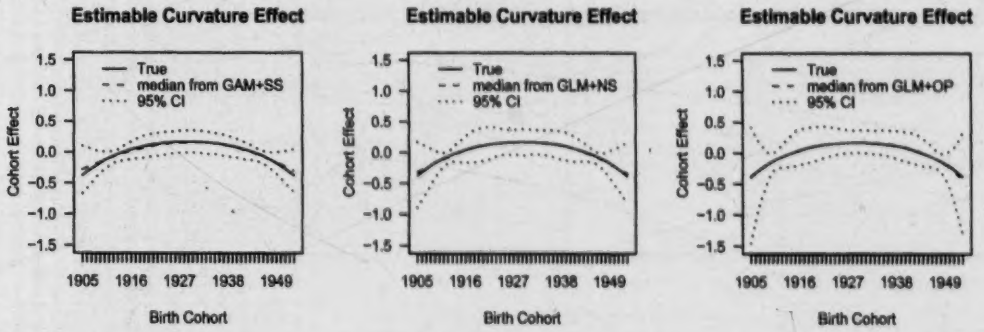


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.14: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	$-.95$	4.8210	4.8425
		GLM+NS	$-.97$	5.9080	5.9182
		GLM+OP	$-.98$	6.9613	6.9638
$\beta_p + \beta_c$	1	GAM+SS	0.96	8.6255	8.6417
		GLM+NS	0.97	9.2077	9.2140
		GLM+OP	0.99	10.0723	10.0739

Figure 4.15: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$.

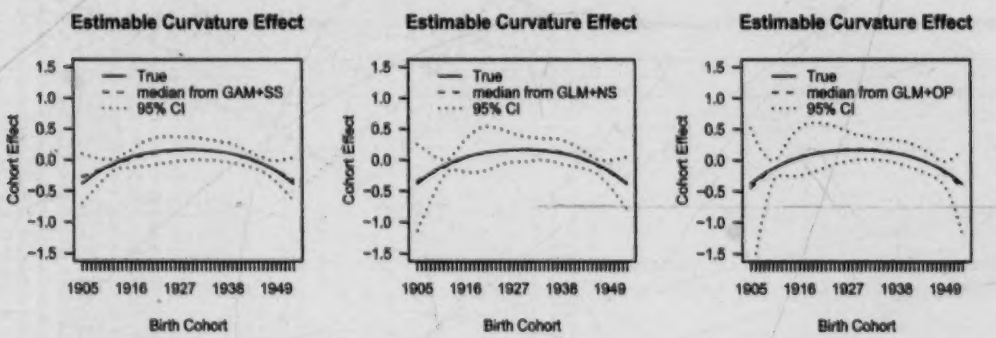


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.15: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-6	GAM+SS	-5.99	5.4467	5.4471
		GLM+NS	-6.11	7.9300	8.0415
		GLM+OP	-6.17	9.8163	10.1085
$\beta_p + \beta_c$	1	GAM+SS	0.94	9.3965	9.4309
		GLM+NS	1.06	12.8803	11.1455
		GLM+OP	1.11	10.0723	13.0113

Figure 4.16: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve II with $\beta_c = 0.01$.

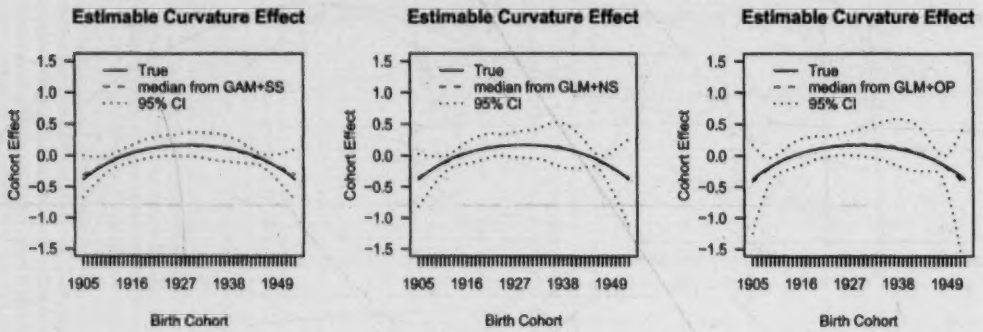


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.16: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	4	GAM+SS	3.99	5.4123	5.4129
		GLM+NS	4.10	7.6297	7.7273
		GLM+OP	4.16	9.4434	9.7011
$\beta_p + \beta_c$	-4	GAM+SS	-4.03	9.2912	9.2997
		GLM+NS	-4.14	10.8405	11.0243
		GLM+OP	-4.19	12.2548	12.6212

Figure 4.17: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

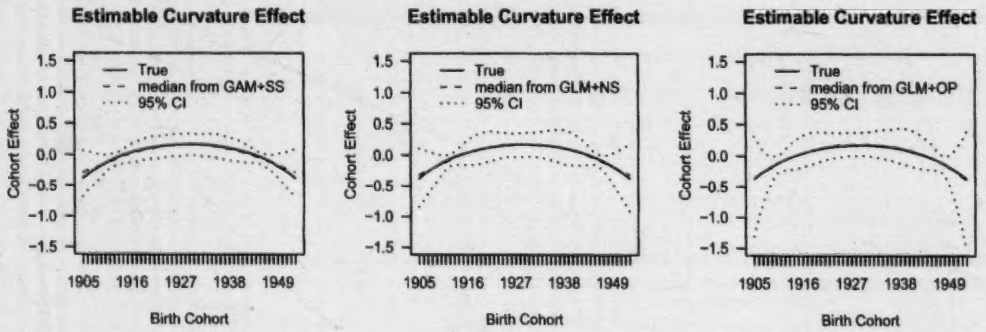


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.17: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-1	GAM+SS	-1.02	4.6466	4.6526
		GLM+NS	-1.00	5.8311	5.8310
		GLM+OP	-0.98	7.0622	7.0653
$\beta_p + \beta_c$	-4	GAM+SS	-4.01	8.3213	8.3225
		GLM+NS	-4.03	8.8857	8.8947
		GLM+OP	-4.05	9.9712	9.9958

Figure 4.18: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

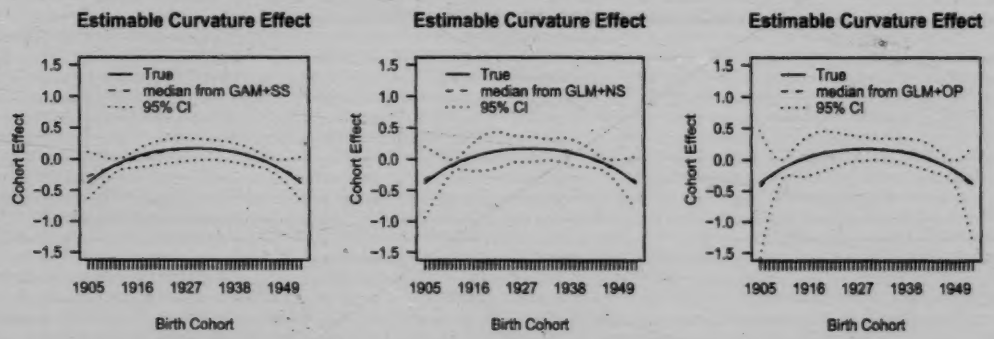


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.18: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
		GAM+SS	-6.03	4.7861	4.7940
$\beta_a - \beta_c$	-6	GLM+NS	-6.08	6.1475	6.2150
		GLM+OP	-6.12	7.3651	7.5065
		GAM+SS	-4.07	8.4798	8.5285
$\beta_p + \beta_c$	-4	GLM+NS	-4.00	9.2271	9.2273
		GLM+OP	-3.98	10.2855	10.2895

Figure 4.19: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve II with $\beta_c = 0.01$.

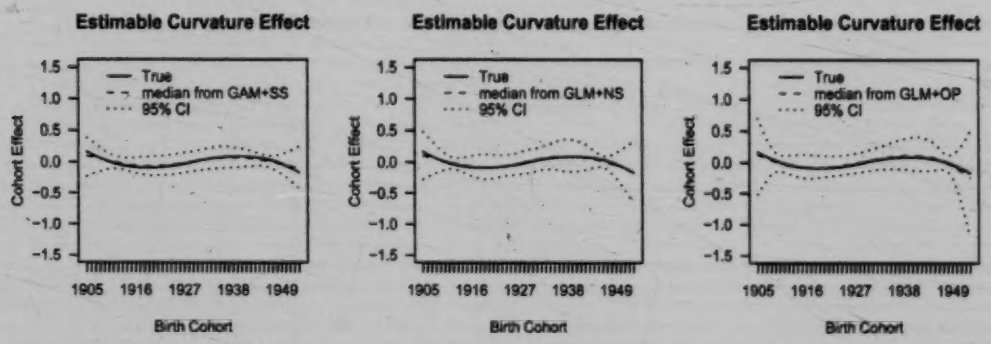


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.19: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	5.587	GAM+SS	5.4926	3.7187	3.8078
		GLM+NS	5.6022	4.4908	4.4931
		GLM+OP	5.6832	5.1859	5.2785
$\beta_p + \beta_c$	4.413	GAM+SS	4.5034	7.1026	7.1844
		GLM+NS	4.4424	7.4320	7.4406
		GLM+OP	4.3455	7.9383	7.9838

Figure 4.20: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = 5.587 \times 10^{-2}$.

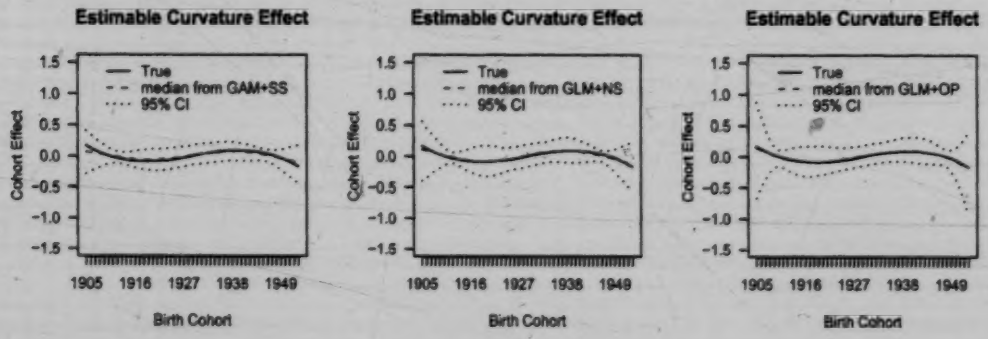


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.20: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	0.587	GAM+SS	0.4497	4.2572	4.4456
		GLM+NS	0.5252	4.8936	4.9318
		GLM+OP	0.5783	5.4040	5.4047
$\beta_p + \beta_c$	4.413	GAM+SS	4.5170	7.5668	7.6741
		GLM+NS	4.4850	7.6964	7.7482
		GLM+OP	4.4157	8.1797	8.1798

Figure 4.21: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$.

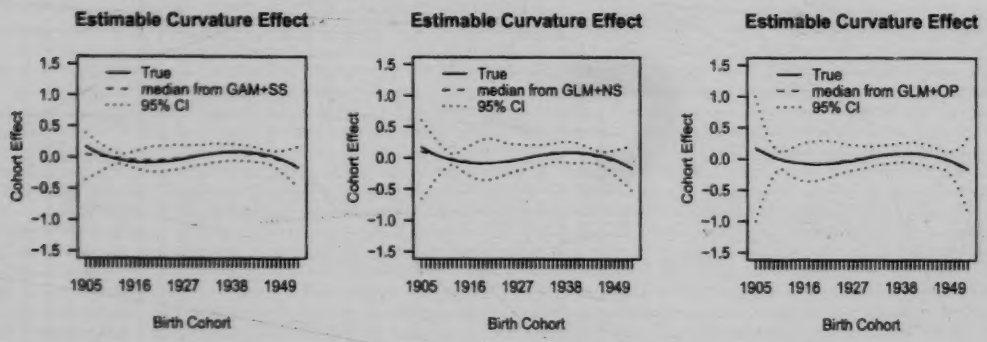


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.21: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-4.413	GAM+SS	-4.6071	4.7749	5.1518
		GLM+NS	-4.5552	6.1468	6.3491
		GLM+OP	-4.5234	6.8972	7.0192
$\beta_p + \beta_c$	4.413	GAM+SS	4.5888	8.1135	8.4225
		GLM+NS	4.5246	8.8046	9.0613
		GLM+OP	4.3455	9.4491	9.5735

Figure 4.22: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.05$ and using birth cohort curve III with $\beta_c = -0.00587$.

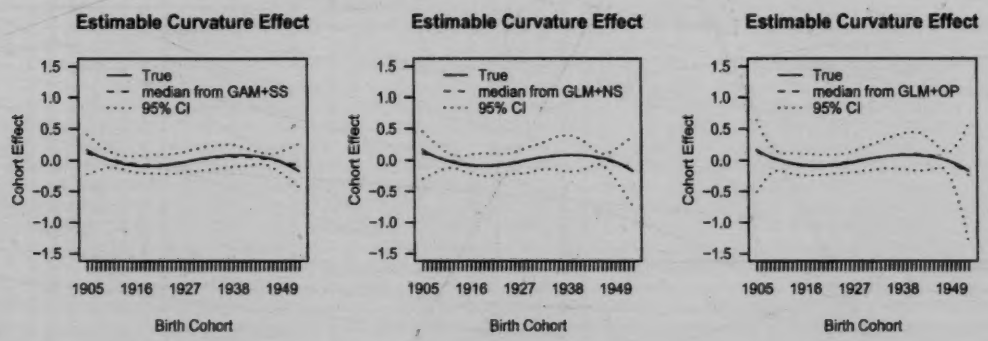


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.22: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
		GAM+SS	5.4737	4.0334	4.1618
$\beta_a - \beta_c$	5.587	GLM+NS	5.5929	5.3028	5.3031
		GLM+OP	5.6910	6.3292	6.4374
		GAM+SS	0.5051	7.9029	7.9700
$\beta_p + \beta_c$	0.587	GLM+NS	0.5795	8.7400	8.7407
		GLM+OP	0.6913	9.7738	9.8826

Figure 4.23: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$.

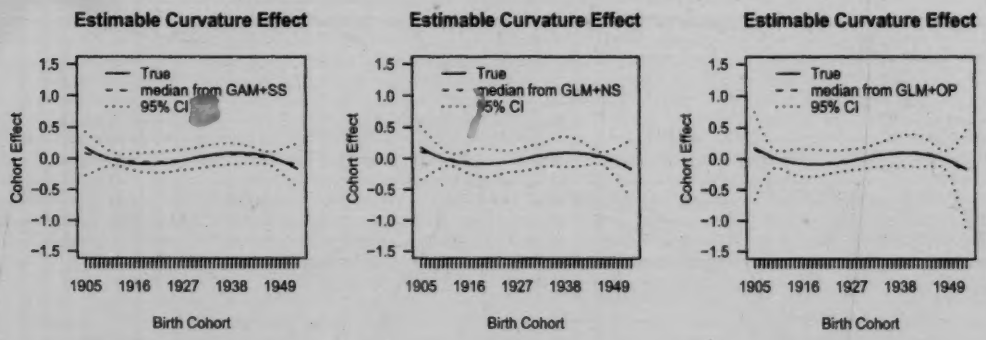


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.23: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	0.587	GAM+SS	0.4755	4.4454	4.5697
		GLM+NS	0.5598	5.1273	5.1347
		GLM+OP	0.6285	5.8968	5.9140
$\beta_p + \beta_c$	-0.587	GAM+SS	-0.4746	7.8380	7.9644
		GLM+NS	-0.5157	8.0112	8.0620
		GLM+OP	-0.6009	8.4927	8.4945

Figure 4.24: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$.

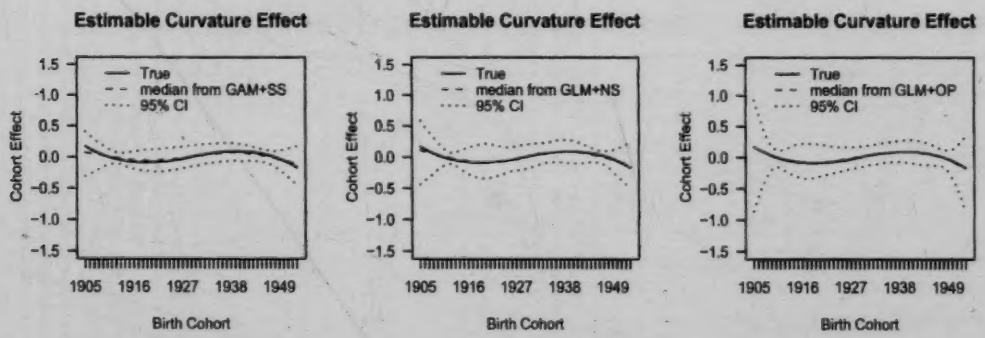


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.24: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	-4.413	GAM+SS	-4.5364	4.5672	4.7195
		GLM+NS	-4.4744	5.3311	5.3688
		GLM+OP	-4.4326	5.8606	5.8645
$\beta_p + \beta_c$	-0.587	GAM+SS	-0.5126	7.8218	7.8771
		GLM+NS	-0.5372	8.1403	8.1652
		GLM+OP	-0.5984	8.5593	8.5606

Figure 4.25: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = 0.00$ and using birth cohort curve III with $\beta_c = -0.00587$.

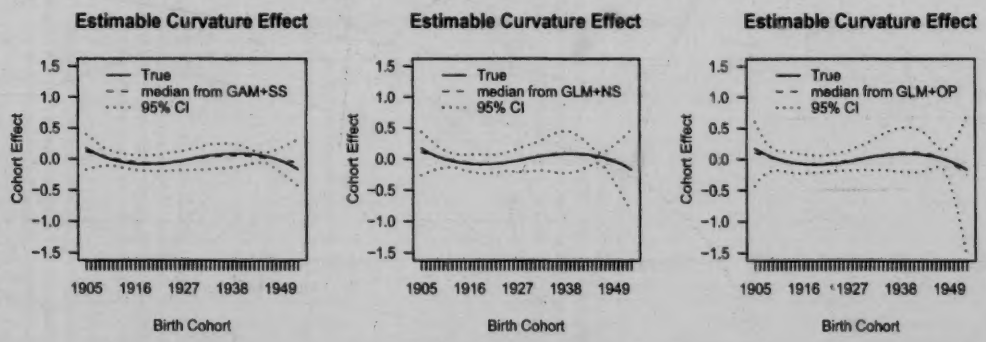


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.25: Estimation results from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	5.587	GAM+SS	5.4888	4.7602	4.8568
		GLM+NS	5.6376	6.9626	6.9882
		GLM+OP	5.7577	8.6850	8.9765
$\beta_p + \beta_c$	−5.587	GAM+SS	−5.4722	8.5388	8.6706
		GLM+NS	−5.5795	10.0148	10.0154
		GLM+OP	−5.7113	11.4504	11.6048

Figure 4.26: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$.

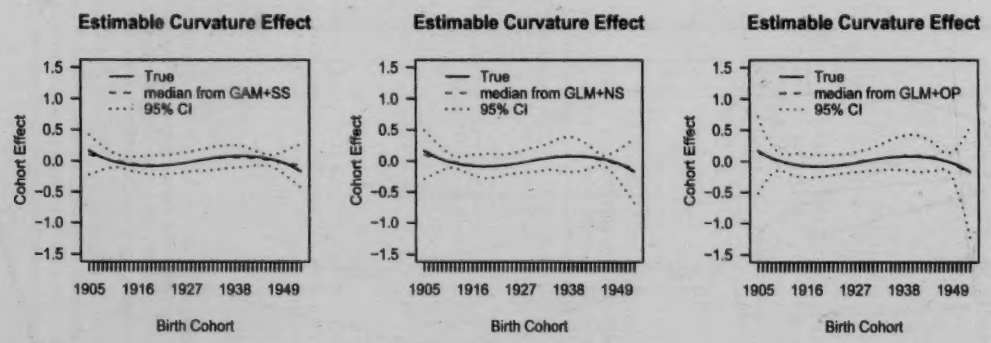


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.26: Estimation results from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
$\beta_a - \beta_c$	0.587	GAM+SS	0.5100	4.0311	4.0905
		GLM+NS	0.6196	5.0787	5.0894
		GLM+OP	0.7047	6.1382	6.2768
$\beta_p + \beta_c$	−5.587	GAM+SS	−5.4793	7.6008	7.7167
		GLM+NS	−5.5431	8.1322	8.1515
		GLM+OP	−5.6471	9.0287	9.0648

Figure 4.27: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = 0.00$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$.

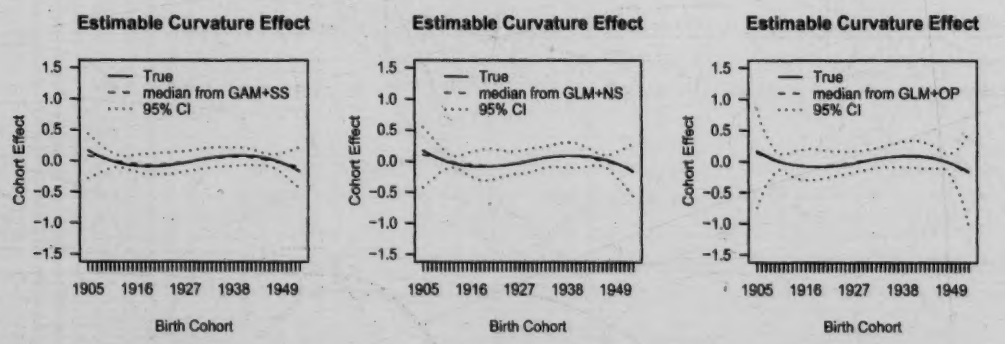


Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.27: Estimation results from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$

Estimable functions	Theoretical Value ($\times 10^{-2}$)	Model	Empirical Mean of estimates ($\times 10^{-2}$)	Empirical Var of estimates ($\times 10^{-5}$)	Empirical MSE of estimates ($\times 10^{-5}$)
		GAM+SS	−4.5700	4.0412	4.2878
$\beta_a - \beta_c$	−4.413	GLM+NS	−4.4880	4.6789	4.7352
		GLM+OP	−4.4317	5.2600	5.2635
		GAM+SS	−5.5003	7.2239	7.2992
$\beta_p + \beta_c$	−5.587	GLM+NS	−5.5368	7.3431	7.3684
		GLM+OP	−5.6146	7.7350	7.7426

Figure 4.28: Estimated curvature effects of birth cohort from the simulation study when $\beta_a = -0.05$, $\beta_p = -0.05$ and using birth cohort curve III with $\beta_c = -0.00587$.



Note: solid curve=constructed curvature effects of cohort; dashed curve=median and dotted curves=2.5 and 97.5 percentiles of the estimates from 1000 simulation runs.

Table 4.28: Analysis of deviance for the mesothelioma data in Alberta.

Models	N[a]	Df	Deviance	$\hat{\phi}^b$
one-year intervals				
GAM+SS	600	638.27	590	1.082
GLM+NS	600	641.24	590	1.087
GLM+OP	600	642.53	590	1.089
five-year intervals				
GLM+F	24	8	7.50	0.938

[a]=number of cells in the Lexis diagram.
[b]=estimates of the dispersion parameter, i.e., $\hat{\phi}$ = deviance/df.

4.2 Analysis of Mesothelioma Data in Alberta

We now analyze the mesothelioma data in Alberta using the three APC smooth models– GAM+SS, GLM+NS and GLM+OP as well as the traditional APC model, which treats the three factors as categorical variable for the data in five-year age and period intervals and will be referred to as GLM+F. The incidence rates considered are in one-year age intervals from 51 to 80 and calender years from 1985 to 2004. To apply the traditional APC model GLM+F, the mesothelioma data is further aggregated into five-year age groups ($51-55$, $56-60$ through $76-80$) and period (i.e. year of diagnosis) groups (1985-1989 up to $2000-2004$); the corresponding ten-year cohort intervals would be $1904-1914$, $1909-1919$ through $1944-1954$, which will be referred to by their mid-points, 1909, 1914 through 1949. For comparison, we decided to use the same number of degrees of freedoms for age, period and cohort when fitting three different smooth models. The degrees of freedom of 3, 3 and 4 for age, period and cohort respectively were found to give the best fits to all the three models simultaneously. Actually, different numbers of degrees of freedom around the values chosen above will not make the model fittings significantly different. When fitting GLM+F the degrees of freedom for each factor is determined by the number of groups upon data aggregation.

The results of goodness of fits are summarized in Table 4.28. All four models seem to give satisfactory fits with the estimates of dispersion parameters being close enough to 1. When forcing the linear trend of cohort effects to be zero, we get the estimates of age slopes and period slopes (i.e., $\beta_a - \beta_c$ and $\beta_p + \beta_c$), which can be interpreted as cross-sectional age trend and overall time trend. Since GAM+SS is a nonparametric model fitting technique, the

Table 4.29: Results for the estimable linear trends in the mesothelioma incidence.

Model	$\beta_a - \beta_c$			$\beta_p + \beta_c$		
	Estimate	SD	P(> |z|)	Estimate	SD	P(> |z|)
One-year interval						
GAM+SS	0.0513			0.0265		
GLM+NS	0.0529	0.0083	< 0.001	0.0246	0.0100	0.0142
GLM+OP	0.0548	0.0088	< 0.001	0.0221	0.0103	0.0328
Five-year interval						
GLM+F	0.2411	0.0391	< 0.001	0.1465	0.0497	0.0185

Table 4.30: Chi-square tests for the estimable curvature effects for mesothelioma data.

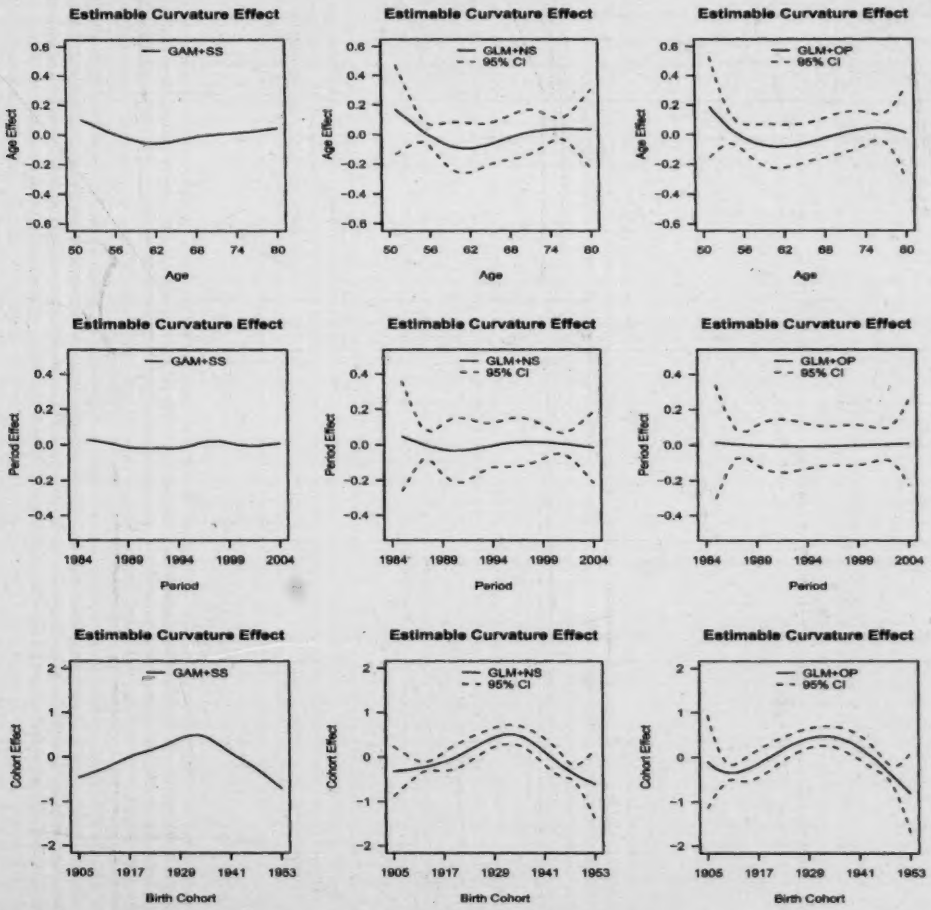
Model	Age			Period			Cohort		
	Df	Chisq	P(Chisq)	Df	Chisq	P(Chisq)	Df	Chisq	P(Chisq)
One-year interval									
GLM+NS	2	3.83	0.15	2	0.39	0.82	3	20.04	< 0.001
GLM+OP	2	4.04	0.13	2	0.52	0.77	3	20.98	< 0.001
Five-year interval									
GLM+F	4	6.81	< 0.12	2	1.62	0.42	7	16.61	0.013

calculation of standard errors of estimable functions is not straightforward and merits future research. All three smooth models using one-year intervals returned similar results while the factor model using five-year groupings returned higher estimates, which is mostly due to the data aggregation. Figure 4.29 shows the three estimable curvature effects estimated using the three smoothing techniques. The plot of the curvature effects from the factor model is shown in Figure 4.30. All four models show us almost the same pattern of effects for each factor.

The estimated cross-sectional age slope and curvature effects suggest that the risk of mesothelioma is increasing as people get older and yet those between 60 and 65 years of age are slightly less likely to have mesothelioma while both the magnitudes and significance tests of curvature effects for age in table 4.30 suggest this pattern is not significant. The estimates of the overall time trend also indicate that the mesothelioma incidence in Alberta is significantly increasing from 1985 to 2004. Although all four plots of curvature effects for period in figure 4.29 showed a slightly concave and then convex pattern, the magnitudes of these effects are so small that the curvature patterns are not significant. Since curvature effects in the trends could be related to exterior changes, such as new methods of diagnosis or disease classification, we could consider excluding the term representing the curvature effect of period from the model if no additional information explained the curvature patterns or changes in the trends. The birth cohort, on the other hand, showed significant curvature effects; the incidence rate increased

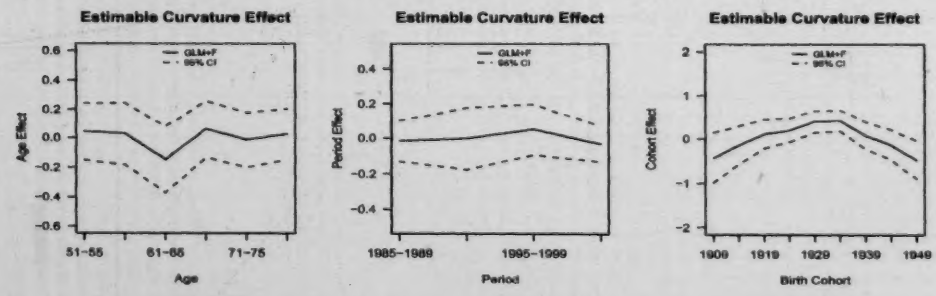
rapidly for early cohorts born from 1905 to 1930 and then decreased sharply for those born afterwards.

Figure 4.29: Estimates of curvature effects from three smooth APC models using 3, 3 and 4 Dfs for Age, Period and Cohort respectively.



Note: dashed curves represent $\tilde{f} \pm 1.96\text{SD}[\tilde{f}]$ from GLM+NS and GLM+OP.

Figure 4.30: Estimates of curvature effects from the APC factor model with 5, 4 and 8 Dfs for Age, Period and Cohort respectively.



Note: dashed curves represent $\tilde{f} \pm 1.96\text{SD}[\tilde{f}]$ from GLM+F.

Chapter 5

Conclusion and Discussion

Age-Period-Cohort model has been widely used in analyzing the time trends disease incidence and mortality rates. In this paper, we are mainly focused on two general categories of Age-Period-Cohort models–smooth APC models using one-year interval and factorial APC models using five-year intervals. The data considered in the factorial APC models are usually tabulated in a two-way contingency table by age group and period (Lexis diagram), which are most often done in five-year intervals. The aggregated incidence counts in each cell of the Lexis diagram are assumed to follow Poisson distributions and the incidence rates in each cell are assumed to be constant. The three factors contribute additive effects to the log rates.

The factorial APC model has been considered as the classic and traditional model for the Age-Period-Cohort analysis, but in some sense, it causes loss of information due to data aggregation, which is important for factorial design to get stable and smoothed effect estimates. However, grouping the data into five-year or ten-year intervals is not important in the framework of smooth APC models, which can accommodate yearly ungrouped data and still fit smoothed curve estimates and it is no longer necessary to make the assumption of constant rates for five or ten years.

The existing smoothing techniques that have been applied to the APC modeling include orthogonal polynomials, which is basically the idea of Holford's parametrization (1983) and have been the most commonly used and reliable solution to the non-identifiability problem, and regression splines by Heuer (1997). In this thesis, we considered the use of another popular smoothing techniques in the APC modeling; we proved that the estimable functions obtained by Holford (1983) also hold in the framework of APC modeling using smoothing splines. In simulation studies we compare the performances of the existing smoothing techniques

GLM+NS, GLM+OP and GAM+SS in the situation where the incidence is very sparse and have fairly large portion of zeroes. We found that GAM+SS is more stable and hence reliable for sparse data compared to the other two, although they are also nice alternatives for this data.

However, due to the nonparametric features of smoothing splines, the calculation of standard deviations for estimable functions is not straightforward and will need study further. When applying either orthogonal polynomials or regression splines, the design matrix can be formulated for each factor. The forecasting can be easily made by extrapolating the estimable functions and hence estimable. Although extrapolation is always risky, it is nevertheless a useful tool for health services planning and prevention strategies. Theoretically, forecasting using GAM+SS is also estimable and possible, and will be explored in future research.

Bibliography

[1] Buja, A., Hastie, T. and Tibshirani, R. Linear Smoothers and Additive Models. Annals of Statistics, 17:453-510, 1989.

[2] Carstensen, B. and Keiding, N. Age-period-cohort Models: Statistical Inference in the Lexis Diagram. Lecture Notes. Department of Biostatistics, University of Copenhagen, 2004 (http://www.biostat.ku.dk/bxc/APC/notes.pdf).

[3] Clayton, D. and Schifflers, E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Statistics in Medicine, 6:449-467, 1987.

[4] Clayton, D. and Schifflers, E. Models for temporal variation in cancer rates. II: Age-periodcohort models. Statistics in Medicine, 6:469-481, 1987.

[5] Dyba, T. and Hakulinen, T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. Statistics in Medicine, 19:1741-1752, 2000.

[6] Dyba, T., Hakulinen, T. and Päivärinta, L. A simple non-linear model in incidence prediction. Statistics in Medicine, 16:2297-2309, 1997.

[7] Fienberg, S. E. and Mason, W. M. (1979). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. In Sociological Methodology, K. F. Schuessler (ed.), 1-67. San Francisco: Josey-Bass.

[8] Hastie, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models. London: Chapman and Hall.

[9] Hastie, T. J., Tibshirani, R. J. and Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag.

[10] Heuer, C. Modelling of time trends and interactions in vital rates using restricted regression splines. Biometrics, 53(1):161-177, 1997.

[11] Hodgson, J.T., McElvenny, D.M., Darnton, A.J., Price, M.J. and Peto, J. The expected burden of mesothelioma mortality in Great Britain from 2002 to 2050. British Journal of Cancer, 92:587-593, 2005.

[12] Holford, T. R. Approaches to fitting age-period-cohort models with unequal intervals. Statistics in Medicine, 25:977-993, 2006.

[13] Holford, T. R. Understanding the effects of age, period, and cohort on incidence and mortality rates. Annual Reviews of Public Health, 12:425-457, 1991.

[14] Holford, T. R. The estimation of age, period and cohort effects for vital rates. Biometrics, 39:311-324, 1983.

[15] Ilg, A.G., Bignon, J. and Valleron, A.J. Estimation of the past and future burden of mortality from mesothelioma in France. Occupational and Environmental Medicine, 55:760-765, 2005.

[16] Kupper, L. L., Janis, J. M., Karmous, A., Greenberg, B. G. Statistical age-period-cohort analysis: A review and critique. Journal of Chronic Disease, 38:811-830, 1985.

[17] Kupper, L. L., Janis, J. M., Salama, I. A., Yoshizawa, C. N., Greenberg, B. G. Age-period-cohort analysis: An illustration of the problems in assessing interaction in one observation per cell data. Communications in Statistics–Theory and Method, 12:2779-2807, 1983.

[18] Møller, B., Fekjær, H., Hakulinen, T., Sigvaldason, H., Storm, H. H., Talbäck, M. and Haldorsen, T. Prediction of cancer incidence in the Nordic contries: empirical comparison of different approaches. Statistics in Medicine, 22:2751-2766, 2003.

[19] Osmond, C. Using age, period and cohort models to estimate future mortality rates. International Journal of Epidemiology, 14(1):124-129, 1985.

[20] Peto, J., Decarli, A., La Vecchia, C., Levi, F. and Negri., E. The European mesothelioma epidemic. British Journal of Cancer, 79:666-672, 1999.

[21] Pelucchi, C., Malvezzi, M., La Vecchia, C., Levi, F., Decarli, A. and Negri, E. The Mesothelioma epidemic in Western Europe: an update. British Journal of Cancer, 90:1022-1024, 2004.

[22] Robertson, C. and Boyle, P. Age-period-cohort analysis of chronic disease rates. I: Modelling approach. Statistics in Medicine, 17:1305-1323, 1998.

[23] Segura, O., Burdorf, A. and Looman, C. Update of predictions of mortality from pleural mesothelioma in the Netherlands. Occupational and Environmental Medicine, 60:50-55, 2003.

[24] Zheng, T., Holford, T. R., Boyle, P., Chen, Y., Ward, B. A., Flannery, J. and Mayne, S. T. Time trend and the age period cohort effect on the incidence of histological types of lung cancer in Connecticut, 1960 to 1989. Cancer, 74:1556-1567, 1994.